



Universidade de Aveiro Departamento de Matemática
2014

**Helena Isabel
Coutinho Galante**

**As variáveis da empregabilidade: uma análise
exploratória de dados**



Universidade de Aveiro Departamento de Matemática
2014

**Helena Isabel
Coutinho Galante**

**As variáveis da empregabilidade: uma análise
exploratória de dados**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Matemática e Aplicações, realizada sob a orientação científica da Doutora Nélia Silva Professora Auxiliar do Departamento de Matemática da Universidade de Aveiro e co orientação da Doutora Isabel Maria Simões Pereira, Professora Auxiliar do Departamento de Matemática da Universidade de Aveiro.

o júri

presidente

Prof. Doutor Agostinho Miguel Mendes Agra
Professor Auxiliar da Universidade de Aveiro.

vogais

Prof. Doutora Helena Maria Pereira Pinto Dourado e Alvéolos
Professora Auxiliar da Universidade de Aveiro.

Prof. Doutora Nélia Maria Marques da Silva
Professora Auxiliar da Universidade de Aveiro (orientadora).

Agradecimentos

Findo este percurso, resta-me agradecer a todas as pessoas que de forma direta ou indireta me apoiaram e que tornaram possível a concretização deste objetivo.

De forma mais particular, agradeço à minha orientadora, Professora Doutora Nélia Silva, e à Professora Doutora Isabel Pereira, coordenadora deste mestrado, pela dedicação, preocupação e simpatia com que sempre me trataram, dando sempre sábios conselhos ao longo deste ano de trabalho.

Agradeço ao meu marido pela paciência, compreensão e apoio incondicional e aos meus filhos por todos os sorrisos, mesmo naqueles momentos em que foram privados da minha atenção.

Por último, uma enorme gratidão aos meus pais que, pelo seu exemplo de vida, me dão sempre força e motivação para lutar pelos objetivos e por me fazerem acreditar que o sonho comanda a vida!

palavras-chave

Análise de dados, Diplomados, Empregabilidade, Tabelas de contingência, Regressão logística.

resumo

O objetivo principal deste trabalho consiste em identificar as variáveis que poderão estar relacionadas com a empregabilidade de diplomados de uma Instituição do Ensino Superior. Para isso recorreu-se à análise das respostas a um questionário efetuado a diplomados dessa Instituição.

Numa primeira fase, os resultados são efetuados através de análises a tabelas de contingência, portanto uma análise essencialmente descritiva, servindo de *input* à fase seguinte, com o intuito de construir um modelo que permita descrever o fenómeno da empregabilidade dos diplomados da Instituição através da análise de regressão logística.

keywords

Data analysis, Graduates, Employability, Contingency tables, Logistic regression.

abstract

The main aim of this work is to show how variables can be related to employability through the results of a survey which was applied to the graduates of an higher education institution. First the results were analyzed using contingency tables, in a mainly descriptive analysis, and its results were the input to the next phase, in order to achieve a model that may describe the graduates employability through the logistic regression analysis.

Índice

| | |
|--|-----------|
| 1. Introdução..... | 1 |
| 1.1. Contextualização e objetivos | 1 |
| 1.2. Estrutura da tese | 2 |
| 2. Análise de Tabelas de Contingência | 4 |
| 2.1. Introdução | 4 |
| 2.2. Noções gerais | 4 |
| 2.2.1. Tabelas de contingência com duas margens livres | 7 |
| 2.2.2. Tabelas de contingência com uma margem fixa | 8 |
| 2.2.3. Análise de resíduos | 10 |
| 2.2.4. Restrições na aplicação ao teste do qui-quadrado | 10 |
| 2.2.5. Medidas de associação para variáveis qualitativas | 11 |
| 3. Regressão Logística | 13 |
| 3.1. Introdução | 13 |
| 3.2. Noções básicas | 14 |
| 3.3. Regressão logística Simples | 15 |
| 3.3.1. Ajustamento do modelo | 15 |
| 3.4. Regressão logística múltipla..... | 17 |
| 3.4.1. Ajustamento do Modelo | 17 |
| 3.5. Interpretação dos parâmetros estimados | 20 |
| 3.6. Construção do modelo logístico | 23 |
| 3.6.1. Escolha do método de seleção de variáveis..... | 27 |
| 4. Análise de Dados | 29 |
| 4.1. Introdução | 29 |
| 4.2. Caracterização sociodemográfica dos respondentes | 30 |
| 4.2.1. Caracterização global | 30 |
| 4.2.2. Caracterização por ciclo de estudos..... | 32 |
| 4.3. Percurso Académico | 36 |
| 4.4. Percurso Profissional..... | 42 |
| 4.4.1. Primeiro Emprego..... | 42 |
| 4.4.2. Diplomados empregados..... | 45 |
| 4.4.3. Diplomados desempregados | 50 |
| 4.5. Modelo logístico da empregabilidade | 53 |
| 5. Conclusão..... | 59 |
| Bibliografia | 62 |
| Anexo: Questionário..... | 64 |

Lista de Tabelas

| | |
|--|----|
| Tabela 2.1 Exemplo de tabela contingência..... | 5 |
| Tabela 2.2 Forma genérica de uma tabela de contingência bidimensional | 5 |
| Tabela 2.3 Forma genérica das probabilidades de uma tabela de contingência com uma margem fixa..... | 8 |
| Tabela 4.1 Situação profissional dos diplomados, por ciclos de estudo..... | 32 |
| Tabela 4.2 Situação profissional dos diplomados, por ciclos de estudo..... | 37 |
| Tabela 4.3 Situação ocupacional no último ano de curso, por ciclos de estudo..... | 39 |
| Tabela 4.4 Continuação de estudos no ensino superior, por ciclo | 41 |
| Tabela 4.5 Fatores que influenciaram a continuação dos estudos, por ciclo de estudos | 41 |
| Tabela 4.6 Motivos para não prosseguir os estudos no ensino superior, por ciclo de estudos | 42 |
| Tabela 4.7 Situação profissional após conclusão do curso, por ciclo..... | 44 |
| Tabela 4.8 Área do 1º emprego após a conclusão do curso, por ciclo de estudos | 45 |
| Tabela 4.9 Área Condição laboral dos diplomados empregados, por ciclos de estudos..... | 46 |
| Tabela 4.10 Área Condição laboral dos diplomados empregados, por ciclos de estudos..... | 48 |
| Tabela 4.11 Razões das dificuldades na obtenção de emprego | 52 |
| Tabela 4.15 Valores das variáveis explicativas no modelo de regressão logística..... | 57 |

Lista de Gráficos

| | |
|--|----|
| Gráfico 4.1 Proporção de respondentes, por ano letivo dos diplomados..... | 29 |
| Gráfico 4.2 Proporção de respondentes, por ciclo de estudos. | 30 |
| Gráfico 4.3 Caixa de bigodes relativo à variável idade dos respondentes | 30 |
| Gráfico 4.4 Distribuição dos respondentes por área científica do curso | 31 |
| Gráfico 4.5 Distribuição dos respondentes de licenciaturas por ano letivo. | 33 |
| Gráfico 4.6 Distribuição dos respondentes de licenciaturas de acordo com a situação profissional, por ano letivo..... | 33 |
| Gráfico 4.7 Distribuição dos respondentes de mestrados integrados, por ano letivo..... | 34 |
| Gráfico 4.8 Distribuição dos respondentes de mestrados integrados de acordo com a situação profissional, por ano letivo..... | 34 |
| Gráfico 4.9 Distribuição dos respondentes de mestrados por ano letivo. | 35 |
| Gráfico 4.10 Distribuição dos respondentes de mestrados de acordo com a situação profissional, por ano letivo..... | 36 |
| Gráfico 4.11 Participação em programas de intercâmbio por ano letivo..... | 38 |
| Gráfico 4.12 Distribuição dos respondentes de mestrados de acordo com a situação profissional, por ano letivo..... | 40 |
| Gráfico 4.13 Situação profissional após a conclusão do curso | 43 |
| Gráfico 4.14 Condição laboral dos diplomados empregados | 45 |
| Gráfico 4.15 Condição laboral dos diplomados empregados, por ano letivo. | 47 |
| Gráfico 4.16 Diplomados empregados segundo a área científica do curso | 49 |
| Gráfico 4.17 Diplomados empregados segundo a área científica do curso, por ano letivo..... | 50 |
| Gráfico 4.18 Diplomados desempregados segundo a área científica do curso | 51 |
| Gráfico 4.19 Diplomados desempregados segundo a área científica do curso e ano letivo. | 52 |

1. Introdução

1.1. Contextualização e objetivos

O desemprego está entre os problemas que mais preocupam não só os cidadãos, mas também os nossos governantes. Segundo o Instituto Nacional de Estatística (INE), a taxa de desemprego em Portugal em 2013 foi de 16.2%. Comumente, ouve-se dizer que este é um problema que afeta toda a sociedade, desde analfabetos a diplomados do Ensino Superior. Mas os dados do INE revelam que o desemprego é menor entre a população que completou um curso de nível superior. Enquanto que a percentagem de desempregados entre a população com curso de ensino superior foi de 12.6%, nos restantes níveis escolaridade as taxas variam entre 17% do nível de ensino básico e 17.4% do ensino secundário ou pós-secundário. Estes dados reforçam a importância da prossecução de estudos após o ensino secundário. No entanto, tendo em consideração as recomendações da *European University Association* no estudo “Portuguese Higher Education: A view from the outside”, e também a atual conjuntura económica, revela-se necessária uma reorganização da oferta formativa do ensino superior. A acrescer a estes factos tem-se que os números do desemprego ou emprego entre os diplomados são também determinantes nas avaliações dos respetivos ciclos de estudos no ensino superior.

Este estudo não procura dar a receita para a obtenção de emprego, mas procura sim responder à questão: quais as variáveis da empregabilidade?

O trabalho surge com a necessidade de analisar a empregabilidade dos diplomados de uma Instituição de Ensino Superior conhecendo o seu percurso académico. Assim, através, do conjunto de respostas a um questionário, que se encontra em Anexo, efetuado aos diplomados de três anos letivos, tenta-se perceber se existem variáveis que propiciam o emprego/desemprego. Para este propósito, foram analisadas um conjunto de variáveis, quer sociodemográficas, quer relacionadas com o percurso académico dos diplomados.

O questionário foi criado e aplicado pelo Serviço de Gestão Académica da Instituição do Ensino Superior. Todos os diplomados em 2008/2009, 2009/2010 e 2010/2011 foram inquiridos

através da plataforma informática existente, durante o mês de novembro de 2012. No total foram notificados 10598 diplomados dos três anos letivos, excluindo da análise 960 diplomados por motivos de ausência de contactos na plataforma.

Neste estudo foram apenas considerados os diplomados de 1º e 2º ciclos de estudo, uma vez que a taxa de resposta dos diplomados de 3º ciclo foi bastante reduzida.

A análise de dados foi efetuada com recurso ao programa *IBM SPSS Statistics 21*.

1.2. Estrutura da tese

Nesta dissertação procura-se não só perceber como se relacionam as variáveis sociodemográficas com o emprego/desemprego mas também como o percurso académico dos diplomados poderá influenciar a sua situação profissional.

Numa primeira fase do trabalho foram utilizadas tabelas de contingência para identificar relações entre as variáveis e a situação profissional. A segunda fase do estudo teve como objetivo encontrar um modelo de regressão que descrevesse a propensão para a empregabilidade. Para isso recorreu-se a uma análise de regressão logística.

A tese está estruturada da seguinte forma:

Nos capítulos 2 e 3 é apresentada a fundamentação teórica dos conceitos que foram utilizados neste trabalho. Assim, no capítulo 2 serão abordadas noções fundamentais sobre tabelas de contingência, analisando-se duas situações: tabelas de contingência com duas margens livres e tabelas de contingência com uma margem fixa. Completa-se este capítulo com a análise de resíduos e apresentando medidas de associação.

O capítulo 3 refere-se à regressão logística. A abordagem a esta técnica é baseada no estudo de Hosmer & Lemeshow (2013). Começa-se por analisar os casos do modelo simples e múltiplo, o significado dos parâmetros estimados e finalmente procede-se à construção do modelo logístico com a definição dos métodos de seleção das variáveis explicativas a considerar no próprio modelo.

A ilustração prática das técnicas abordadas nos capítulos 2 e 3 é feita no capítulo 4 com a sua aplicação aos dados recolhidos resultantes do inquérito aos diplomados de uma Instituição do Ensino Superior. Numa primeira parte pretende-se avaliar como se relacionam as diversas variáveis analisadas com a situação profissional dos diplomados através de tabelas de contingência; portanto uma análise essencialmente descritiva. Na segunda parte deste capítulo recorre-se à regressão logística com o intuito de obter um modelo que explique o comportamento da empregabilidade dos diplomados e identificar as variáveis que propiciam o emprego/desemprego.

O capítulo 5 compreende as principais conclusões deste trabalho, bem como possíveis futuras abordagens ao tema em estudo.

2. Análise de Tabelas de Contingência

2.1. Introdução

Numa população os elementos são descritos por características ou atributos que podem ser de natureza quantitativa ou qualitativa. Os atributos quantitativos podem assumir diferentes quantidades ou valores numéricos (por exemplo, idade, altura, distância). No que concerne a atributos qualitativos, estes assumem categorias. Se não existir uma ordem associada às categorias (por exemplo, o género) diz-se que são provenientes de uma escala nominal. Se, por outro lado existir uma ordenação nos valores, então são medidas numa escala ordinal (por exemplo, o nível de escolaridade). Segundo Everit (1997) as variáveis podem ser classificadas segundo uma classificação dicotómica (as variáveis assumem um de dois valores) ou de acordo com uma classificação múltipla (é mais comum, onde as variáveis podem assumir mais de dois valores).

No presente estudo dá-se evidência às variáveis qualitativas, ou variáveis categóricas¹, onde os diplomados são descritos por um conjunto de variáveis categóricas nomeadamente pelo ciclo de estudos (1º ciclo, 2º ciclo integrado ou 2º ciclo) ou situação profissional (empregado ou desempregado).

2.2. Noções gerais

As relações entre variáveis categóricas são usualmente organizadas em tabelas que Karl Pearson apelidou de tabelas de contingência ou de classificação cruzada. Estas tabelas mostram o relacionamento existente entre as variáveis através da respetiva distribuição conjunta (Agresti, 2002).

Considere-se o seguinte exemplo expresso na Tabela 2.1, relativo à situação profissional (desempregado ou empregado) dos diplomados que concluíram o curso nos anos letivos 2008/2009, 2009/2010 e 2010/2011. Trata-se de uma tabela de contingência que relaciona as variáveis “situação profissional” com o “ano de conclusão do curso” dos diplomados. As células

¹ Qualquer variável quantitativa pode ser transformada numa variável categórica.

da tabela correspondem a frequências ou contagens. Assim, por exemplo, o total de desempregados que concluíram o curso em 2008/2009 é de 205.

| | 2008/2009 | 2009/2010 | 2010/2011 | Total |
|--------------|-----------|-----------|-----------|-------|
| Desempregado | 205 | 397 | 810 | 1412 |
| Empregado | 397 | 484 | 497 | 1378 |
| Total | 602 | 881 | 1307 | 2790 |

Tabela 2.1 Exemplo de tabela contingência.

Seguidamente apresenta-se a forma genérica de uma tabela de contingência. Para isso considere-se X e Y duas características da população que se subdividem, respetivamente, em r e c categorias. Onde, x_i representa a categoria i ($i=1, \dots, r$) da variável X e y_j a categoria j ($j=1, \dots, c$) da variável Y . Seja $n \geq 1$ o total de observações de uma amostra casual a classificar. A Tabela 2.2 representa a tabela de contingência $r \times c$, tendo como variáveis X e Y com r e c categorias, respetivamente.

| $\begin{matrix} Y \\ X \end{matrix}$ | y_1 | ... | y_j | ... | y_c | Totais Linha (n_{i+}) |
|--------------------------------------|--|-----|--|-----|--|---|
| x_1 | n_{11} | | n_{1j} | | n_{1c} | Total marginal de X na categoria 1 |
| x_i | n_{i1} | ... | n_{ij} | ... | n_{ic} | Total marginal de X na categoria i |
| x_r | n_{r1} | | n_{rj} | | n_{rc} | Total marginal de X na categoria r |
| Total Coluna (n_{+j}) | Total marginal de Y na categoria 1 | ... | Total marginal de Y na categoria j | ... | Total marginal de Y na categoria c | Total global da amostra |

Tabela 2.2 Forma genérica de uma tabela de contingência bidimensional.

Cada elemento n_{ij} ($i=1, \dots, r$ e $j=1, \dots, c$) corresponde ao número de observações classificadas simultaneamente como pertencentes à categoria i da variável X e à categoria j da variável Y . Considera-se que as categorias são mutuamente exclusivas.

O total de observações na i -ésima categoria da variável X é denotado por n_{i+} . Da mesma forma, n_{+j} representa a soma das observações da j -ésima categoria da variável Y. Estes valores são denominados como totais marginais e são definidos da seguinte forma:

$$n_{i+} = n_{i1} + n_{i2} + \dots + n_{ic} = \sum_{j=1}^c n_{ij} \quad n_{+j} = n_{1j} + n_{2j} + \dots + n_{rj} = \sum_{i=1}^r n_{ij}.$$

Desta forma,

$$n = n_{++} = \sum_{i=1}^r \sum_{j=1}^c n_{ij}.$$

corresponde ao número de observações dos dados.

[Voltando ao exemplo anterior (Tabela 2.1), tem-se uma tabela de contingência 2×3 onde $n = 2790$ e os totais marginais são: $n_{1+} = 1412$, $n_{2+} = 1378$, $n_{+1} = 602$, $n_{+2} = 881$ e $n_{+3} = 1307$.]

Considere-se que a probabilidade de um elemento pertencer à categoria i da variável X e simultaneamente à categoria j da variável Y é denotada por p_{ij} ($i=1,\dots,r$ e $j=1,\dots,c$). De forma análoga, p_{i+} e p_{+j} denotam as probabilidades de uma observação pertencer, respetivamente, à categoria, i e j . Naturalmente tem-se que:

$$p_{i+} = \sum_{j=1}^c p_{ij}, \quad p_{+j} = \sum_{i=1}^r p_{ij} \quad e \quad \sum_{i=1}^r \sum_{j=1}^c p_{ij} = 1.$$

O objetivo principal da análise de tabelas de contingência é averiguar se as variáveis aleatórias que presidiram à classificação cruzada são ou não independentes. Caso não exista independência entre as variáveis aleatórias, conclui-se que existe associação entre elas. Os coeficientes de associação, que se abordarão numa secção posterior, permitem quantificar o grau da associação entre variáveis.

As margens da tabela de contingência (os números totais das observações) podem ser livres, fixas ou ambas as situações (uma margem fixa e outra livre). Seguidamente serão abordadas as duas situações mais comuns: tabelas com margens livres e tabelas com uma das margens fixa.

2.2.1. Tabelas de contingência com duas margens livres

Seguindo a terminologia anterior, considere-se que, na população de onde a amostra é extraída, a probabilidade de uma observação pertencer à categoria i da variável X e simultaneamente à categoria j da variável Y é p_{ij} ($P(X=i, Y=j)$). A distribuição conjunta do vetor aleatório constituído pelo número de observações nas várias células, N_{ij} , $i=1, \dots, r$; $j=1, \dots, c$, considerando fixa a dimensão da amostra, é multinomial de parâmetros (n, p_{ij}) , cuja função de probabilidade é:

$$P(N_{11} = n_{11}, N_{12} = n_{12}, \dots, N_{rc} = n_{rc}) = \frac{n!}{\prod_{i=1, \dots, r} \prod_{j=1, \dots, c} n_{ij}!} \prod_{i=1, \dots, r} \prod_{j=1, \dots, c} p_{ij}^{n_{ij}}.$$

Uma vez que se desconhece a distribuição de probabilidades, p_{i+} e p_{+j} , estas são estimadas através das frequências observadas. Segundo Everit (1997, p.6) $\hat{p}_{i+} = \frac{n_{i+}}{n}$ ($i = 1, \dots, r$) e $\hat{p}_{+j} = \frac{n_{+j}}{n}$ ($j=1, \dots, c$) constituem as melhores estimativas para os valores procurados e são obtidos através do método da máxima verosimilhança.

Considerando que as variáveis X e Y são independentes tem-se:

$$\hat{p}_{ij} = \hat{p}_{i+} \hat{p}_{+j} = \frac{n_{i+}}{n} \frac{n_{+j}}{n}, \quad i = 1, \dots, r \text{ e } j = 1, \dots, c.$$

Permitindo estimar as frequências esperadas na ij -ésima célula da tabela, $E_{ij} = np_{ij}$. Resultando,

$$\hat{E}_{ij} = e_{ij} = n \hat{p}_{i+} \hat{p}_{+j} = n \frac{n_{i+}}{n} \frac{n_{+j}}{n} = \frac{n_{i+} n_{+j}}{n} \quad (i = 1, \dots, r \text{ e } j = 1, \dots, c).$$

Testar a independência entre as variáveis X e Y equivale a testar

$$H_0: p_{ij} = p_{i+} \cdot p_{+j} \quad \text{versus} \quad H_1: \exists (i,j), i \neq j : p_{ij} \neq p_{i+} \cdot p_{+j}.$$

Uma vez que as frequências observadas têm distribuição binomial e supondo que H_0 é verdadeira, a estatística de teste é definida com base na diferença entre os valores observados (n_{ij}) e os valores estimados (e_{ij}) da seguinte forma:

$$X_{rc}^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - e_{ij})^2}{e_{ij}}.$$

A estatística X_{rc}^2 aproxima-se de uma distribuição assintótica do qui-quadrado com $(r - 1)(c - 1)$ graus de liberdade (Cramér, 1946; Agresti, 2002). O valor desta função depende essencialmente da diferença entre as frequências observadas e esperadas $(n_{ij} - e_{ij})$, logo será menor no caso das variáveis serem independentes, e terá um valor relativamente grande caso contrário. Assim, considerando o *p-value* definido por:

$$p - value = P(X_{rc}^2 > k \mid H_0 \text{ verdadeira}).$$

sendo k o valor observado da estatística de teste, rejeita-se a hipótese nula, ao nível de significância α , se $p - value < \alpha$, para um dado valor α inicialmente fixado (valores usuais: 0.05, 0.01 ou 0.1).

2.2.2. Tabelas de contingência com uma margem fixa

Considere-se agora uma tabela idêntica à anterior, mas onde apenas uma das margens é fixa. Sem perda de generalidade considere-se que os totais marginais das linhas são fixos, isto é, n_{i+} são fixos para $i=1, \dots, r$. Isto corresponde a ter r amostras, uma em cada linha da tabela, obtidas de modo independente e de dimensões iguais, respetivamente, a $n_{1+}, n_{2+}, \dots, n_{r+}$.

Deste modo, a probabilidade de um indivíduo pertencer à categoria j de Y , estando na categoria i de X , é dada por:

$$p_{j|i} = P(Y = j \mid X = i), \quad j = 1, \dots, c; i = 1, \dots, r.$$

Resultando a tabela

| $\begin{matrix} Y \\ X \end{matrix}$ | y_1 | ... | y_j | ... | y_c | Totais Linha |
|--------------------------------------|-----------|-----|-----------|-----|-----------|-----------------|
| x_1 | $p_{1 1}$ | | $p_{j 1}$ | | $p_{c 1}$ | 1 |
| | | | | | | ... |
| x_i | $p_{1 i}$ | ... | $p_{j i}$ | ... | $p_{c i}$ | 1 |
| | ... | | ... | | ... | ... |
| x_r | $p_{1 r}$ | | $p_{j r}$ | | $p_{c r}$ | 1 |
| Totais coluna | p_{+1} | ... | p_{+j} | ... | p_{+c} | 1 |

Tabela 2.3 Forma genérica das probabilidades de uma tabela de contingência com uma margem fixa.

Nesta situação a hipótese a testar não é a independência mas sim a homogeneidade entre as variáveis. Em termos simbólicos pretende-se testar:

$$H_0: p_{j|1} = p_{j|2} = \dots = p_{j|r}, j = 1, \dots, c \text{ versus } H_1: \exists_{i,k} : p_{j|i} \neq p_{j|k}.$$

Em termos de distribuição, sabe-se que a distribuição conjunta de um vetor aleatório constituído pelo número de observações em cada célula, com n_{i+} fixo, é multinomial de parâmetros n_{i+} e $p_{j|i}$. Assim, a função de probabilidade é definida pela seguinte expressão:

$$P(N_{i1} = n_{i1}, N_{i2} = n_{i2}, \dots, N_{ic} = n_{ic} | n_{i+}) = \frac{n_{i+}!}{\prod_{j=1}^c n_{ij}!} \prod_{j=1}^c p_{j|i}^{n_{ij}}.$$

onde $n_{ij} = 0, \dots, n_{i+}$ e $\sum_{j=1}^c n_{ij} = n_{i+}$.

Uma vez que as r linhas são independentes entre si, a probabilidade de se obter o conjunto das observações n_{ij} da tabela, conhecidos os n_{i+} e as probabilidades $p_{j|i}$ é dada por

$$P(N_{i1} = n_{i1}, N_{i2} = n_{i2}, \dots, N_{rc} = n_{rc} | n_{1+}, n_{2+}, \dots, n_{r+}) = \prod_{i=1}^r \left(\frac{n_{i+}!}{\prod_{j=1}^c n_{ij}!} p_{j|i}^{n_{ij}} \prod_{j=1}^c p_{j|i}^{n_{ij}} \right).$$

Sendo H_0 verdadeira tem-se

$$e_{ij} = \hat{E}_{j|i} = n_{i+} \hat{p}_{j|i} = \frac{n_{i+} \cdot n_{+j}}{n}, i = 1, \dots, r \text{ e } j = 1, \dots, c.$$

e a estatística de teste é dada por

$$X_{rc}^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - e_{ij})^2}{e_{ij}}.$$

que tem distribuição assintótica do qui-quadrado com $(r - 1)(c - 1)$ graus de liberdade.

Assim, considerando o p -value definido por:

$$p - \text{value} = P(X_{rc}^2 > k | H_0 \text{ verdadeira}).$$

rejeita-se a hipótese nula, ao nível de significância α , se $P - \text{value} < \alpha$.

Constatando-se que o teste de homogeneidade, para grandes amostras sob este modelo, é idêntico ao teste utilizado para testar a independência, embora em contextos diferentes.

2.2.3. Análise de resíduos

A análise de resíduos permite conhecer quais as categorias que mais contribuem para o valor da estatística do qui-quadrado através da análise dos resíduos normalizados:

$$r_{ij} = \frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}}}.$$

Sendo a estimativa da sua variância dada por

$$v_{ij} = (1 - \frac{n_{i+}}{n})(1 - \frac{n_{+j}}{n}).$$

resulta que o resíduo padronizado de cada célula da tabela de contingência é dado por:

$$d_{ij} = \frac{r_{ij}}{\sqrt{v_{ij}}}.$$

Quando as variáveis são independentes, d_{ij} segue uma distribuição aproximadamente normal reduzida. Assim, para se saber quais os resíduos significativos comparam-se d_{ij} com um quantil de probabilidade $1 - \alpha/2$ da distribuição normal reduzida. No caso de $\alpha=0.05$, consideram-se significativos valores $|d_{ij}|$ maiores que 1.96, que corresponde ao valor do quantil de probabilidade igual a 0.975.

2.2.4. Restrições na aplicação ao teste do qui-quadrado

Existem algumas limitações no recurso ao teste do qui-quadrado no âmbito do estudo das relações entre as variáveis de uma tabela de contingência.

Normalmente, em tabelas do tipo 2 x 2, define-se que os valores das frequências esperadas não devem ser inferiores a 10 e em tabelas de dimensão maior os valores de E_{ij} não devem ser menores que 5. Nos casos em que se verifiquem frequências esperadas abaixo destes valores mínimos, é usual fundir-se categorias de modo a serem satisfeitas as condições de aplicação do teste do qui-quadrado [1].

2.2.5. Medidas de associação para variáveis qualitativas

O teste do qui-quadrado indica se existe ou não relação entre as variáveis envolvidas. Porém, no caso de haver relacionamento entre variáveis, este teste nada diz quanto à intensidade. Neste contexto, poderá recorrer-se às medidas de associação.

As medidas de associação mais adequadas para variáveis qualitativas, em que pelo menos uma das variáveis é nominal, são: o coeficiente de contingência de Pearson, o coeficiente ϕ , o coeficiente de Tschuprow, ou o coeficiente de Cramér, que se baseiam na estatística de teste do qui-quadrado utilizada nos testes de independência.

- Coeficiente de Contingência de Pearson

$$P = \sqrt{\frac{\chi^2}{\chi^2 + n}}.$$

- Coeficiente ϕ (apenas para tabelas do tipo 2 x 2)

$$\phi = \sqrt{\frac{\chi^2}{n}}.$$

- Coeficiente de Tschuprow

$$T = \sqrt{\frac{\phi^2}{\sqrt{(r-1)(c-1)}}} = \sqrt{\frac{\chi^2}{n\sqrt{(r-1)(c-1)}}}.$$

- Coeficiente de Cramér

$$V = \sqrt{\frac{\phi^2}{\min(c-1, r-1)}} = \sqrt{\frac{\chi^2/n}{\min(c-1, r-1)}}.$$

Alguns destes coeficientes apresentam algumas limitações de utilização. Por exemplo, o coeficiente ϕ é definido para tabelas do tipo 2 x 2 e segundo Kendal, referido por Goodman & Kruskal (1954), os coeficientes P e T, foram construídos numa tentativa de normalizar χ^2 de modo a situarem-se entre 0 e 1 e assumir os valores extremos. No entanto, o coeficiente de contingência de Pearson nunca atinge o valor 1 e o seu valor máximo depende do número de linhas e colunas, já o coeficiente de Tschuprow só atinge o valor 1 apenas se o número de linhas for igual ao número de colunas. Por sua vez, o coeficiente de Cramér V não depende da dimensão nem da forma da tabela. Os seus valores situam-se entre 0, que significa a ausência

de associação (portanto, independência) entre as variáveis, e 1, que expressa uma associação perfeita entre as variáveis.

Posto isto, no presente estudo optou-se pelo coeficiente de Cramér para medir a associação entre variáveis.

3. Regressão Logística

3.1. Introdução

Em estatística a análise de regressão consiste numa técnica exploratória que procura exprimir o relacionamento entre variáveis, quando este existe, através de uma função. Conhecendo essa função é possível explicar e prever fenômenos.

A aplicação da regressão logística tem sido recorrente nas várias áreas de estudo, tanto numa perspectiva de análise de relações entre variáveis, como numa perspectiva preditiva. Por exemplo, na educação, Saha (2011) recorreu a um modelo logístico para avaliar quais as variáveis que poderiam influenciar os resultados curriculares dos estudantes de Tripura (Índia). Já na área da gestão financeira, Missuni *et al* (2002) construíram um modelo logístico constituído por cinco variáveis capaz de prever situações de (in)solvência. Ferreira (2013) apresenta uma aplicação de modelos de regressão logística na área da medicina dentária. Também na área da saúde, Bonellie (2011) aplicou uma análise de regressão logística para avaliar quais os fatores que influenciam o peso de recém-nascidos à nascença.

O que distingue os modelos de regressão linear dos modelos de regressão logística é fundamentalmente a variável dependente, que no âmbito logístico assume na maioria das situações apenas dois valores: 0 ou 1, sendo portanto de natureza dicotómica. No entanto, pode também ser de natureza policotómica; mas esta situação não vai estar no âmbito de aplicação deste estudo.

Os modelos de regressão logística seguem os mesmos princípios que os modelos de regressão linear. Assim, as técnicas usadas no âmbito da regressão linear serão análogas às usadas na regressão logística, como se pode ver no trabalho apresentado por Hosmer & Lemeshow (2013).

3.2. Noções básicas

Genericamente, nos modelos de regressão linear pretende-se analisar o comportamento de uma variável dependente Y (variável explicada ou também denominada variável endógena) com base num conjunto de p variáveis independentes X_1, \dots, X_p (variáveis explicativas ou também designadas variáveis exógenas). Assim, procura-se a função que permita traduzir este comportamento:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} + \varepsilon_i, i = 1, \dots, n.$$

onde n corresponde à dimensão da amostra, ε_i é variável aleatória exógena não observada - que expressa o desvio das observações à média condicional - e β_i , $i=0, \dots, p$, representam os parâmetros do modelo (constantes). Assume-se que ε_i tem distribuição normal com valor médio nulo e variância constante.

O valor médio condicional da variável de resposta Y , dado um valor x da variável independente, $E(Y|x)$, constitui um dado fundamental na análise do problema de qualquer modelo de regressão.

Como já se referiu anteriormente na regressão logística a variável dependente Y é dicotómica, ou seja, assume dois valores possíveis, 0 ou 1. Consequentemente, enquanto que nos modelos de regressão linear $E(Y|x)$ pode assumir qualquer valor, no contexto logístico o valor de $E(Y|x)$ varia entre 0 e 1 ($0 \leq E(Y|x) \leq 1$). Neste caso é usual denotar $E(Y|x)$ por $\pi(x)$.

Sendo $y = \pi(x) + \varepsilon$, tem-se que, nos modelos logísticos, ε pode assumir dois valores:

$$\varepsilon = \begin{cases} -\pi(x) & \text{com probabilidade } 1 - \pi(x) & \text{se } y = 0 \\ 1 - \pi(x) & \text{com probabilidade } \pi(x) & \text{se } y = 1 \end{cases}.$$

Assim, ε tem uma distribuição cujo valor médio é nulo e a variância é igual a $\pi(x)[1 - \pi(x)]$. A distribuição condicional da variável resposta $Y|x$ segue uma distribuição binomial com probabilidade de sucesso $\pi(x)$.

À razão $\frac{\pi(x)}{1-\pi(x)}$ dá-se o nome de *Odds*, em alguns estudos também é comum chamar-se possibilidade, vantagem ou *chance*.

Seguidamente e por uma questão de simplicidade, na secção 3.3. irá abordar-se o caso da regressão logística simples, isto é, dependendo de uma única variável independente e na secção 3.4. será então desenvolvida a abordagem múltipla.

3.3. Regressão logística Simples

Quando se tem uma única variável independente, a forma genérica de $\pi(x)$ é expressa por:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}.$$

A linearização de $\pi(x)$ assume um aspeto importante na regressão logística. Esta transformação é designada por transformação *logit* e é definida da seguinte forma:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \ln \left[\frac{\frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}}{1 - \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}} \right] = \ln \left[\frac{\frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}}{\frac{1}{1 + e^{\beta_0 + \beta_1 x}}} \right] = \ln(e^{\beta_0 + \beta_1 x}) = \beta_0 + \beta_1 x.$$

A função *logit* $g(x)$ tem características importantes neste contexto: é linear nos seus parâmetros (β_0 e β_1), pode ser contínua e pode assumir qualquer valor real dependendo da variação da variável independente (Hosmer & Lemeshow, 2013).

3.3.1. Ajustamento do modelo

Considere-se uma amostra constituída por n observações independentes do par (x_i, y_i) , onde $i=1, \dots, n$, y_i denota o valor da variável de resposta (dependente) dicotómica e x_i representa o i -ésimo valor da variável independente. Sendo a variável resposta dicotómica, esta assume dois valores: 0 ou 1, que significam a ausência ou presença da característica em estudo, respetivamente.

O ajustamento do modelo ao conjunto das observações passa pela estimação dos parâmetros β_0 e β_1 de $g(x)$.

Na regressão linear a estimação dos parâmetros é feita com recurso ao método dos mínimos quadrados, que sob as condições usuais apresentam boas propriedades. No entanto, essas propriedades deixam de ser válidas quando se aplica num modelo em que a variável resposta é dicotómica. No caso logístico, uma vez que a variável dependente é dicotómica, o método utilizado é o da máxima verosimilhança. Este método tem como objetivo determinar os valores de $\beta=(\beta_0, \beta_1)$ que maximizam a probabilidade de obter o conjunto dos valores observados. Assim, os estimadores determinados serão aqueles que melhor se adequam aos dados observados (Hosmer & Lemeshow, 2013).

Atendendo a que:

- i) a variável dependente, Y , assume os valores 0 ou 1;
- ii) $P(Y = 1 | x) = \pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$, portanto, $P(Y = 0 | x) = 1 - \pi(x)$,

então, a contribuição de cada par (x_i, y_i) na função de verosimilhança é traduzida na seguinte expressão:

$$\pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \text{ com } y_i \in \{0,1\}.$$

Assumindo que as observações são independentes e identicamente distribuídas, a função de verosimilhança é obtida como o produto dos termos da função de probabilidade.

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} \cdot (1 - \pi(x_i))^{1-y_i}.$$

Logaritimizando $l(\beta)$ obtém-se:

$$L(\beta) = \ln(l(\beta)) = \ln \left[\prod_{i=1}^n \pi(x_i)^{y_i} \cdot (1 - \pi(x_i))^{1-y_i} \right] = \sum_{i=1}^n [y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]].$$

Substituindo $\pi(x_i)$ e $1 - \pi(x_i)$ na expressão resulta que:

$$L(\beta) = \ln(l(\beta)) = \sum_{i=1}^n \left[y_i (\beta_0 + \beta_1 x_i) + \ln \left[\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right] \right] = \sum_{i=1}^n [y_i \beta_0 + y_i \beta_1 x_i - \ln(1 + e^{\beta_0 + \beta_1 x_i})].$$

Para encontrar os valores β que maximizam $L(\beta)$, deriva-se esta função em ordem aos parâmetros e iguala-se seu resultado a zero obtendo-se:

$$\sum [y_i - \pi(x_i)] = 0 \quad \text{e} \quad \sum x_i [y_i - \pi(x_i)] = 0.$$

Estas funções são não-lineares nos parâmetros e portanto requerem métodos iterativos para a sua resolução. McCullagh e Nelder (1989) resolveram estas equações recorrendo a um procedimento iterativo que recorre ao método ponderado dos mínimos quadrados. No entanto, a maioria dos *softwares* estatísticos, nomeadamente o SPSS, utiliza o método de Newton-Raphson (Ferreira, 2013).

3.4. Regressão logística múltipla

Considere-se agora a situação dos modelos de regressão logística múltipla, onde existem mais variáveis independentes, $\mathbf{x} = \mathbf{x}^T = (x_1, x_2, \dots, x_p)$.

Neste caso, $E(Y | \mathbf{x}) = \pi(\mathbf{x})$:

$$\pi(\mathbf{x}) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}.$$

E consequentemente a transformação *logit* da regressão logística múltipla é definida por:

$$g(\mathbf{x}) = \ln \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p.$$

3.4.1. Ajustamento do Modelo

No caso múltiplo, para ajustar o modelo é necessário estimar os parâmetros $\beta^T = (\beta_0, \beta_1, \dots, \beta_p)$. Tal como no caso da regressão logística com uma variável, também no caso múltiplo se recorre ao método da máxima verosimilhança.

Assumindo que as observações são independentes, a função de verosimilhança define-se da seguinte forma:

$$L(\beta) = \ln(l(\beta)) = \sum_{i=1}^n [y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]].$$

De forma análoga ao caso da secção anterior, obtém-se:

$$L(\beta) = \sum_{i=1}^n [y_i \beta_0 + y_i \beta_1 x_{i1} + \dots + y_i \beta_p x_{ip} - \ln(1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}})].$$

Para se obterem os estimadores de máxima verosimilhança ter-se-á de resolver um sistema de $p+1$ equações (de verosimilhança), constituídas pelas derivadas parciais da log verosimilhança relativamente a cada um dos parâmetros. Devido à maior complexidade inerente a esta situação, tal como se fez no modelo de regressão logístico simples, é necessário um *software* adequado para se resolver o sistema de equações referido. Além das estimativas de máxima verosimilhança dos parâmetros, representados por $\hat{\beta}$, podem calcular-se as variâncias e covariâncias dos parâmetros estimados, que se obtêm a partir da matriz constituída pelas segundas derivadas parciais da função log verosimilhança. Estas derivadas parciais têm a forma geral:

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 \pi_i (1 - \pi_i) \quad \text{e} \quad \frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_l} = - \sum_{i=1}^n x_{ij} x_{il} \pi_i (1 - \pi_i),$$

onde $j, l=0, \dots, p$ e $\pi_i = \pi(x_i)$.

Seja $I(\beta)$ a matriz quadrada de dimensão $(p+1)$, denominada Matriz de Informação observada, composta pelos valores simétricos dos termos das últimas duas equações. Através da inversão de $I(\beta)$ obtém-se as variâncias dos coeficientes e as covariâncias entre os coeficientes estimados, que se denotará por $Var(\beta)$ (portanto, igual a $I(\beta)^{-1}$).

Seja $Var(\beta_j)$ o j -ésimo elemento da diagonal principal da matriz, que corresponde à variância de $\hat{\beta}_j$. Considere-se $Cov(\beta_j, \beta_l)$ um elemento arbitrário fora da diagonal da matriz, que corresponderá à covariância de $\hat{\beta}_j$ e de $\hat{\beta}_l$.

Os estimadores das variâncias e covariâncias são obtidos de $I(\beta)^{-1}$ quando se substitui β pelo seu estimador $\hat{\beta}$. $\widehat{Var}(\hat{\beta}_j)$ e $\widehat{Cov}(\hat{\beta}_j, \hat{\beta}_l)$, $j, l=0, \dots, p$ representam assim, os respetivos valores da matriz estimada.

Os desvios padrão dos coeficientes estimados são obtidos através $[\widehat{Var}(\hat{\beta}_j)]^{1/2}$, $j=0, 1, \dots, p$, e serão importantes para testar os coeficientes e estimar de intervalos de confiança.

A formulação da matriz de informação é útil na análise do modelo ajustado. Seja $\hat{I}(\hat{\beta}) = X^T \hat{V} X$, onde X é uma matriz de dimensão $n \times (p+1)$ cuja primeira coluna é constituída por 1's, e as restantes pelos valores referentes aos valores observados das variáveis independentes, e V é uma matriz diagonal de dimensão n constituída pelo elemento genérico $\hat{\pi}_i(1 - \hat{\pi}_i)$, onde $\hat{\pi}_i = \hat{\pi}(x_i)$, isto é:

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \quad \text{e} \quad V = \begin{bmatrix} \hat{\pi}_1(1 - \hat{\pi}_1) & 0 & \dots & 0 \\ 0 & \hat{\pi}_2(1 - \hat{\pi}_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \hat{\pi}_n(1 - \hat{\pi}_n) \end{bmatrix}.$$

Uma vez ajustado o modelo, é necessário avaliar a significância/validade do modelo estimado, de forma a identificar quais as variáveis com maior influência no modelo estimado. Através do teste da razão de verosimilhança (*Likelihood Ratio*) avalia-se a hipótese dos coeficientes β_j , com exceção do termo independente β_0 , serem simultaneamente nulos. Ou seja,

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0 \quad \text{versus} \quad H_1: \exists_j: \beta_j \neq 0, j = 1, \dots, p.$$

Usa-se para estatística de teste

$$D = -2 \ln \left[\frac{\text{Função de máxima verosimilhança do modelo corrente}}{\text{Função de máxima verosimilhança do modelo saturado}} \right]$$

$$= -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right],$$

sendo o modelo saturado aquele que abrange todas as variáveis e interações e o modelo corrente o que contém apenas as variáveis escolhidas para o estudo. Usualmente a estatística D é denominada por *Deviance*.

A significância da estatística do teste da razão de verosimilhanças é testada utilizando a distribuição de χ^2 cujo número de graus de liberdade é igual ao número de variáveis independentes no modelo, não incluindo a constante (Venticinque *et al*, 2007). Rejeitar a hipótese nula significa que pelo menos um dos coeficientes é estatisticamente diferente de zero e por conseguinte há que testar cada um dos coeficientes. Para isso é possível recorrer ao teste de Wald, que averigua se cada coeficiente é significativamente diferente de zero, traduzindo-se na relação entre a respetiva variável independente com a variável dependente.

Assim, a hipótese nula é expressa da seguinte forma:

$$H_0: \beta_j = 0 \text{ onde } j = 0, 1, \dots, p.$$

A estatística de teste é definida por:

$$W_j = \frac{\hat{\beta}_j}{\sqrt{\text{var}(\hat{\beta}_j)}}.$$

Assumindo a hipótese nula verdadeira, esta estatística segue uma distribuição normal reduzida. Como Silva (2011) observa - ver também Hauck e Donner (1977) - o teste Wald pode falhar, pois por vezes rejeita variáveis que são estatisticamente significativas; nesse caso aconselha-se a que os coeficientes das variáveis identificadas como não significativas sejam testados de acordo com o teste da razão de verossimilhança.

3.5. Interpretação dos parâmetros estimados

Para se interpretar os parâmetros estimados do modelo logístico, recorre-se à função de *Odds*. No entanto, considerando que na análise de regressão, a variável dependente pode ser influenciada por variáveis quantitativas e qualitativas é aconselhável proceder a uma transformação das variáveis. As variáveis quantitativas são facilmente convertidas numa escala. No caso das variáveis qualitativas, uma vez que indicam a presença ou a ausência de uma característica ou atributo, são transformadas em variáveis artificiais, comumente designadas por variáveis mudas/auxiliares (*dummies*), que permitem quantificar os atributos assumindo os valores 1 ou 0.

1º Caso: quando a variável independente é dicotómica

Se a variável independente X é dicotómica, assumindo um dos dois valores 0 ou 1, então $\pi(x)$ assumirá apenas os valores $\pi(0)$ e $\pi(1)$, levando a que a distribuição de Y se defina da seguinte forma:

| | $P(Y=y \mid X=0)$ | | $P(Y=y \mid X=1)$ |
|-------|-------------------|-------|-------------------|
| $y=0$ | $1 - \pi(0)$ | $y=0$ | $1 - \pi(1)$ |
| $y=1$ | $\pi(0)$ | $y=1$ | $\pi(1)$ |

onde

$$\pi(0) = \frac{e^{\beta_0}}{1+e^{\beta_0}}; \quad \pi(1) = \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}; \quad 1 - \pi(0) = \frac{1}{1+e^{\beta_0}}; \quad 1 - \pi(1) = \frac{1}{1+e^{\beta_0+\beta_1}}.$$

Quando $X=1$, o que significa a presença da característica, tem-se que o *Odds* é dado por

$$\frac{\pi(1)}{1-\pi(1)}. \text{ Por sua vez, quando } X=0, \text{ o Odds é } \frac{\pi(0)}{1-\pi(0)}.$$

Nestes casos a transformação *logit* traduz-se em:

$$g(1) = \ln \left[\frac{\pi(1)}{1-\pi(1)} \right] = \beta_0 + \beta_1 \quad e \quad g(0) = \ln \left[\frac{\pi(0)}{1-\pi(0)} \right] = \beta_0.$$

A razão de *Odds*, comumente designada por OR (abreviatura do inglês *Odds Ratio*) é expressa por:

$$OR = \frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}}.$$

Aplicando o logaritmo a OR obtém-se:

$$\ln(OR) = \ln \left[\frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}} \right] = g(1) - g(0).$$

A esta diferença dá-se o nome diferença *logit*.

Substituindo os valores de $g(1)$ e $g(0)$ simplifica-se a expressão de OR, obtendo-se:

$$OR = e^{\beta_1}.$$

Para amostra suficientemente grandes, o estimador para β_j , $\hat{\beta}_j$, segue uma distribuição aproximadamente normal reduzida. Assumindo essa aproximação à distribuição normal, encontra-se o intervalo de confiança para β_j definido por:

$$\hat{\beta}_j \pm Z_{1-\frac{\alpha}{2}} \sqrt{\widehat{var}(\hat{\beta}_j)}$$

onde, $Z_{1-\frac{\alpha}{2}}$ representa o quantil de probabilidade $\left(1 - \frac{\alpha}{2}\right)$ da distribuição normal reduzida.

Desta forma o intervalo de confiança para e^{β_1} é dado por:

$$\left[e^{\hat{\beta}_1 - Z_{1-\frac{\alpha}{2}} \sqrt{\widehat{var}(\hat{\beta}_1)}}, e^{\hat{\beta}_1 + Z_{1-\frac{\alpha}{2}} \sqrt{\widehat{var}(\hat{\beta}_1)}} \right].$$

Se o intervalo de confiança contiver o valor 1 - significando que $\beta_1=0$ - conclui-se que não existe relação significativa entre as variáveis X e Y ao nível de confiança de $100(1-\alpha)\%$.

2º Caso: quando a variável independente é policotômica

No caso em que X tem mais de duas categorias, deve-se recorrer a várias variáveis auxiliares. Estas variáveis assumem dois valores possíveis 0 ou 1. Considere-se como exemplo a variável “ciclo de estudos”. Existem três categorias: Licenciaturas, Mestrados e Mestrados Integrados. Neste caso tem-se 2 variáveis *dummies*, D_1 e D_2 que se definem da seguinte forma:

| | D1 | D2 |
|--------------------|----|----|
| Licenciatura | 0 | 0 |
| Mestrado | 1 | 0 |
| Mestrado Integrado | 0 | 1 |

Quando se trata de Licenciatura, D_1 e D_2 são nulas, no caso de Mestrados $D_1=1$ e $D_2=0$ e se for Mestrado Integrado, então $D_1=0$ e $D_2=1$.

Assim, se existem r categorias distintas para a variável X^2 , então existem $r-1$ variáveis *dummies* definidas de forma análoga ao exemplo anterior:

| Categorias de X | D_1 | D_2 | | D_{r-1} |
|-----------------|-------|-------|-----|-----------|
| C_1 | 0 | 0 | ... | 0 |
| C_2 | 1 | 0 | ... | 0 |
| ... | ... | ... | ... | ... |
| C_r | 0 | 0 | ... | 1 |

Neste caso, considera-se uma categoria como referência, as restantes categorias são comparadas à de referência, sendo o cálculo de OR obtido da mesma forma que no 1º caso tal como os intervalos de confiança.

² Por exemplo a área científica do curso que compreende dez categorias possíveis.

3º Caso: quando a variável independente é contínua

Quando a variável independente, X é contínua, a interpretação do coeficiente estimado depende de saber quantas unidades deverá aumentar X , de modo a que a interpretação seja considerada adequada. Assim, há que perceber como se comporta a função *logit*, $g(x) = \beta_0 + \beta_1 x$, com o aumento de unidades. Assumindo que $g(x)$ é linear em x e que se pretende obter um estimador para razão de *odds* de $x+1$ (incremento de uma unidade) *versus* x , uma vez que a função *logit* de $x+1$ é expressa por: $g(x+1) = \beta_0 + \beta_1(x+1)$, tem-se que o valor estimado da diferença de *logit*, $\hat{g}(x+1) - \hat{g}(x)$, é:

$$\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_1 - \hat{\beta}_0 - \hat{\beta}_1 x = \hat{\beta}_1.$$

Donde se conclui que $\widehat{OR} = e^{\hat{\beta}_1}$ é exatamente igual ao valor obtido no 1º caso, quando a variável independente é dicotómica.

Considere-se agora que o valor do incremento é c . Naturalmente tem-se que $\widehat{OR} = e^{c\hat{\beta}_1}$. Desta forma o erro padrão $c\hat{\beta}_1$ é expresso por:

$$\sqrt{\widehat{var}(c\hat{\beta}_1)} = |c| \sqrt{\widehat{var}(\hat{\beta}_1)}.$$

Assim, o intervalo de confiança para $e^{c\hat{\beta}_1}$ ao nível de significância α é dado:

$$\left[e^{c\hat{\beta}_1 - z_{1-\frac{\alpha}{2}} |c| \sqrt{\widehat{var}(\hat{\beta}_1)}}, e^{c\hat{\beta}_1 + z_{1-\frac{\alpha}{2}} |c| \sqrt{\widehat{var}(\hat{\beta}_1)}} \right].$$

Em suma, a interpretação do valor estimado da razão de *odds* é semelhante nos casos da variável independente ser contínua ou nominal (ver Hosmer & Lemeshow, 2013, para mais detalhe).

3.6. Construção do modelo logístico

Hosmer & Lemeshow (2013) definem algumas etapas importantes a considerar na construção de um modelo logístico. Para estes autores a análise inicia-se com a identificação das variáveis

independentes candidatas a entrar no primeiro modelo logístico, ou seja, aquelas que parecem contribuir para o valor da variável resposta. Assim, através de uma análise logística simples, as variáveis independentes são avaliadas individualmente. A escolha das candidatas recai sobre as variáveis cujo nível de significância do teste de Wald seja inferior a 0.25. Hosmer & Lemeshow (2013, p.91) recomendam este valor tendo por base os trabalhos de Bendel e Afifi (1977) e de Mickey e Greenland (1989), onde se demonstra que a escolha de um *p-value*, inferior a 0.05, poderá resultar na não inclusão de variáveis relevantes na análise. Em contrapartida, a opção por um *p-value* superior poderá levar a incluir no modelo inicial algumas variáveis com pouca relevância na análise. Consequentemente, estes autores sugerem que antes de tomar um modelo como definitivo, devem-se analisar criticamente todas as variáveis que foram incluídas.

Tendo identificado as variáveis independentes, constrói-se o modelo múltiplo, avaliando-se a importância, ou o contributo na variável resposta, de cada uma das variáveis independentes através do *p-value* do teste de Wald. Todas as variáveis que não sejam significativas estatisticamente (considerando o nível de significância de 0.05) devem ser retiradas da análise, construindo-se um novo modelo. O modelo resultante deve ser sempre comparado com o anterior (modelo saturado) aplicando-se para isso o teste da razão de verossimilhança.

Este processo de retirar e reajustar o modelo é repetido até se verificar que as variáveis explicativas incluídas no modelo descrevem adequadamente o comportamento da variável resposta. Nesta situação o teste de Hosmer-Lemeshow poderá ser um recurso na avaliação do modelo estimado. Este teste permite avaliar se o modelo encontrado explica bem os dados observados.

Teste de Hosmer-Lemeshow

Considerando X e Y as variáveis do modelo, definidas como anteriormente, Hosmer e Lemeshow (2013, p.157) propuseram duas estratégias de agrupamento destas variáveis tendo por base as probabilidades estimadas. Seja J o número de padrões de covariância (ver exemplo adiante) entre as variáveis, e $J=n$, então as n colunas correspondem aos n valores das probabilidades estimadas, onde a 1ª coluna contém o menor valor e a última (n -ésima)

coluna, o maior valor. As estratégias apresentadas pelos dois autores baseiam-se por um lado no cálculo de percentis das probabilidades estimadas e, por outro, em considerar pontos de corte, ou seja, considerar valores fixos das probabilidades estimadas.

Desta forma, no primeiro método, consideram-se $g = 10$ grupos onde o primeiro contém os $n'_1 = n/10$ subconjuntos com as menores probabilidades estimadas e o último grupo compreende os $n'_{10} = n/10$ subconjuntos com as maiores probabilidades estimadas. Usualmente estes grupos são denominados por “decis de risco”.

No segundo método, os $g=10$ grupos resultam de pontos de corte definidos entre os valores de $k/10, k = 1, 2, \dots, 9$ e de grupos que contém todos os subconjuntos com as probabilidades estimadas entre pontos de corte adjacentes. Por exemplo, o primeiro grupo contém todos os subconjuntos onde as probabilidades estimadas são menores ou iguais a 0.1, enquanto que, o último grupo abarca todos os subconjuntos onde as probabilidades estimadas são maiores que 0.9.

Para qualquer uma destas duas estratégias o teste de Hosmer-Lemeshow é calculado com base na estatística:

$$\hat{C} = \sum_{k=1}^g \left[\frac{(o_{1k} - \hat{e}_{1k})^2}{\hat{e}_{1k}} + \frac{(o_{0k} - \hat{e}_{0k})^2}{\hat{e}_{0k}} \right].$$

Onde:

$$o_{1k} = \sum_{j=1}^{c_k} y_j ;$$

$$o_{0k} = \sum_{j=1}^{c_k} (m_j - y_j);$$

$$\hat{e}_{1k} = \sum_{j=1}^{c_k} m_j \hat{\pi}_j ;$$

$$\hat{e}_{0k} = \sum_{j=1}^{c_k} m_j (1 - \hat{\pi}_j).$$

c_k denota o número padrões de covariação no k-ésimo grupo e m_j o número de observações no j-ésimo padrão de covariância, $j=1,2,...,J$. Para melhor compreender estes conceitos, considere-se o seguinte exemplo retirado de Kislaya (2012).

Exemplo:

Sejam dois conjuntos de dados, que compreendem 11 subconjuntos definidos da seguinte forma:

| Conjunto 1 | | | | | | Conjunto 2 | | | | | |
|------------|---|----------------|----------------|----------------|----------------|------------|---|----------------|----------------|----------------|----------------|
| Subconj | y | x ₁ | x ₂ | x ₃ | x ₄ | Subconj | y | x ₁ | x ₂ | x ₃ | x ₄ |
| 1 | 0 | 1 | 2 | 0 | 1 | 1 | 0 | 1 | 10.2 | 0 | 1 |
| 2 | 0 | 0 | 2 | 0 | 1 | 2 | 0 | 0 | 8.7 | 1 | 0 |
| 3 | 1 | 2 | 3 | 1 | 1 | 3 | 1 | 0 | 8.7 | 1 | 0 |
| 4 | 0 | 2 | 5 | 0 | 0 | 4 | 1 | 0 | 8.7 | 1 | 0 |
| 5 | 0 | 1 | 1 | 1 | 1 | 5 | 1 | 1 | 10.2 | 0 | 1 |
| 6 | 1 | 0 | 7 | 1 | 1 | 6 | 1 | 0 | 8.7 | 1 | 0 |
| 7 | 0 | 1 | 3 | 0 | 1 | 7 | 1 | 1 | 10.2 | 0 | 1 |
| 8 | 1 | 2 | 4 | 0 | 1 | 8 | 0 | 2 | 9.4 | 0 | 1 |
| 9 | 1 | 0 | 7 | 1 | 0 | 9 | 1 | 2 | 9.4 | 0 | 1 |
| 10 | 0 | 0 | 2 | 0 | 1 | 10 | 0 | 1 | 10.2 | 0 | 1 |
| 11 | 1 | 1 | 1 | 0 | 0 | 11 | 1 | 1 | 10.2 | 0 | 1 |

É possível verificar que no primeiro conjunto todas as 11 linhas (os 11 subconjuntos das quatro variáveis explicativas) são diferentes, então neste caso, o número de padrões de covariância, J , é igual a n (ou seja, 11), e o número de observações entre os padrões de covariância é $m_1 = m_2 = \dots = m_{11} = 1$. Adicionalmente, neste caso, o total de sucessos ($y=1$) é 5 e o total de insucessos ($y=0$) é 6.

Por sua vez, no segundo conjunto, existem apenas 3 subconjuntos diferentes: {1,5,7,10,11}; {2,3,4,6}; e {8,9}. Neste caso, J é 3 e o número de observações entre os 3 padrões de covariância são respetivamente $m_1 = 5$, $m_2 = 4$ e $m_3 = 2$. Neste caso, o total de sucessos é $y_1 = 3 + 3 + 1 = 7$ e o total de insucessos $y_0 = 2 + 1 + 1 = 4$.

Voltando à estatística \hat{C} , Hosmer & Lemeshow (2013, p.158) referem que, com alguma álgebra, é possível obter a seguinte relação:

$$\hat{C} = \sum_{k=1}^g \left[\frac{(O_{1k} - n'_{1k} \bar{\pi}_k)^2}{n'_{1k} \bar{\pi}_k (1 - \bar{\pi}_k)} \right].$$

Onde $\bar{\pi}_k$ denota a média da probabilidade estimada do k-ésimo grupo, ou seja,

$$\bar{\pi}_k = \frac{1}{n'_{1k}} \sum_{j=1}^{c_k} m_j \hat{\pi}_j.$$

Estes autores demonstraram no seu trabalho de 1980 que quando $J=n$ e o modelo ajustado é o correto, a estatística do teste de Hosmer-Lemeshow, \hat{C} , segue uma distribuição χ^2_{g-2} . Esta constatação é de resto corroborada pelo trabalho de Canary (2012) referenciado em Hosmer-Lemeshow (2013, p.159).

Na estratégia de agrupamento inicial, estes autores defendem que o método onde resultam os “decis de grupo” apresenta melhor aderência à distribuição χ^2_{g-2} que o método baseado nos pontos de corte, especialmente em situações onde as probabilidades estimadas são inferiores a 0.2. A grande parte dos *softwares* estatísticos (incluindo o SPSS) permitem obter o valor de \hat{C} e o respetivo *p-value* considerando 10 grupos (decis de grupo).

3.6.1. Escolha do método de seleção de variáveis

A abordagem tradicional para a construção de um modelo estatístico envolve a procura do melhor modelo que explique, sem perda significativa, a variação da variável resposta. A inclusão de todas as variáveis relacionadas no modelo dificulta a interpretação e a generalização. O fundamento para esta abordagem é o de fornecer um modelo tão completo quanto possível dentro de cada conjunto de dados, pois variáveis que individualmente não são representativas, em conjunto podem ter considerável representatividade. Assim, devem ser seguidos os passos definidos anteriormente para a construção de um modelo de regressão logístico tendo em consideração a seguinte informação.

Existem três métodos para se proceder à inclusão das variáveis explicativas no modelo: *enter*, *forward*, *backward* e *stepwise*. Estes métodos distinguem na forma como “escolhem” as variáveis a incluir no modelo.

No método *enter*, as variáveis explicativas são incluídas no modelo todas em bloco. O método *forward* inicia-se com o modelo mais simples, sem qualquer variável explicativa, e em cada interação deste método adicionam-se as variáveis até se obter o modelo saturado. A primeira variável a ser incluída no modelo é aquela que apresentar o maior valor de correlação com a variável resposta. Por sua vez, o método *backward* atua de modo contrário, inicia-se com o modelo saturado até obter-se o modelo simples. A primeira variável a ser removida por este método é aquela que apresenta menor correlação com a variável resposta. A estimação dos parâmetros nos métodos *forward* e *backward* pode ser feita através de vários critérios dos quais se destacam a razão de verossimilhanças e o teste de Wald. Existe ainda a combinação dos métodos *backward* e *forward*, resultado no *stepwise*.

4. Análise de Dados

4.1. Introdução

De modo a conhecer o percurso académico e profissional dos diplomados de uma Instituição do Ensino Superior, foi efetuado um inquérito aos diplomados de três anos letivos: 2008/2009, 2009/2010 e 2010/2011. Nesta inquirição procurou-se conhecer não só a empregabilidade entre os diplomados, mas também perceber se existe alguma relação entre o percurso académico e o percurso profissional dos mesmos.

Foram inquiridos 10655 diplomados obtendo-se 2882 respostas, ou seja, uma taxa de resposta de 27%. O Gráfico 4.1 mostra que é entre os diplomados mais recentes que se obtém o maior número de respostas. Apenas 22.2% dos respondentes são diplomados de 2008/2009, por sua vez, 46.3% dos respondentes diplomaram-se em 2010/2011.

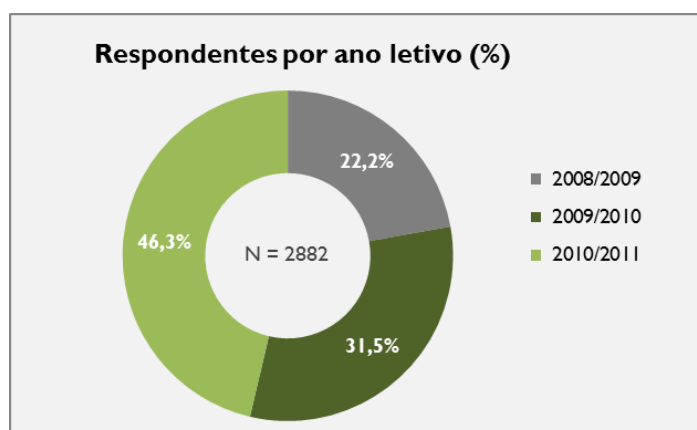


Gráfico 4.1 Proporção de respondentes, por ano letivo dos diplomados.

No que concerne à distribuição dos respondentes por ciclos de estudo - Licenciaturas (1º ciclo), Mestrados Integrados (2º ciclo – MI) e Mestrados³ (2º ciclo) – verifica-se que 54,5% dos respondentes são diplomados de licenciaturas e os restantes 45.5%, dividem-se de forma análoga pelos mestrados, integrados e não integrados (Gráfico 4.2).

³ Ao longo do estudo designação de “mestrados” refere-se aos mestrados não integrados.

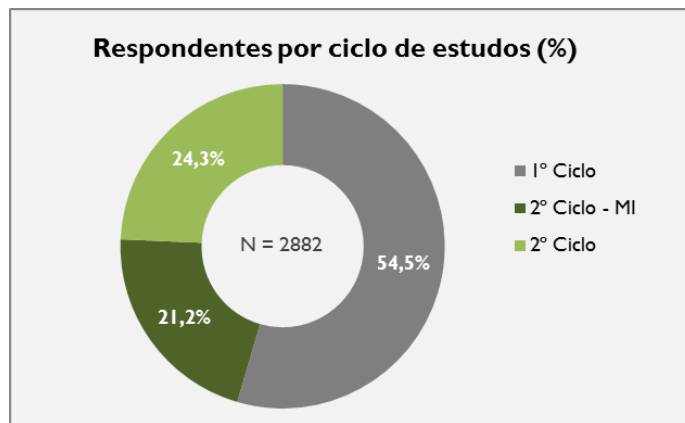


Gráfico 4.2 Proporção de respondentes, por ciclo de estudos.

4.2. Caracterização sociodemográfica dos respondentes

4.2.1. Caracterização global

No que concerne à caracterização da população respondente, 57.3% são do sexo feminino e a média etária ronda os 28 anos (Desvio Padrão, $s=5.5$). Como se constata pelo Gráfico 4.3, o diagrama do tipo caixa de bigodes da variável idade, os valores dos três quartis são, respetivamente, 25, 27 e 29 anos. A idade de 90% dos respondentes não ultrapassa os 35 anos.

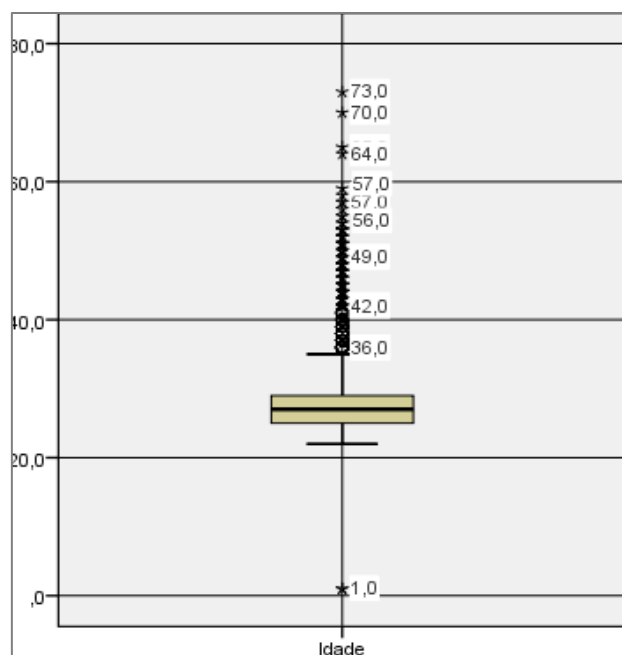


Gráfico 4.3 Caixa de bigodes relativo à variável idade dos respondentes.

A maioria dos respondentes é solteira (87,2%) e provêm maioritariamente (cerca de 78%) da região Centro. Apenas se encontram 65 diplomados estrangeiros, onde se destaca o Brasil como país estrangeiro mais representado entre os respondentes (22).

Os cursos dos 2882 diplomados respondentes dividem-se de acordo com o Gráfico 4.4 pelas diversas áreas científicas⁴. As áreas menos representadas são as Ciências do Desporto e Arquitetura e Construção (não excedem os 4%). Por outro lado, a área com maior proporção de respondentes, ultrapassando os 15%, é a das Ciências Empresarias.

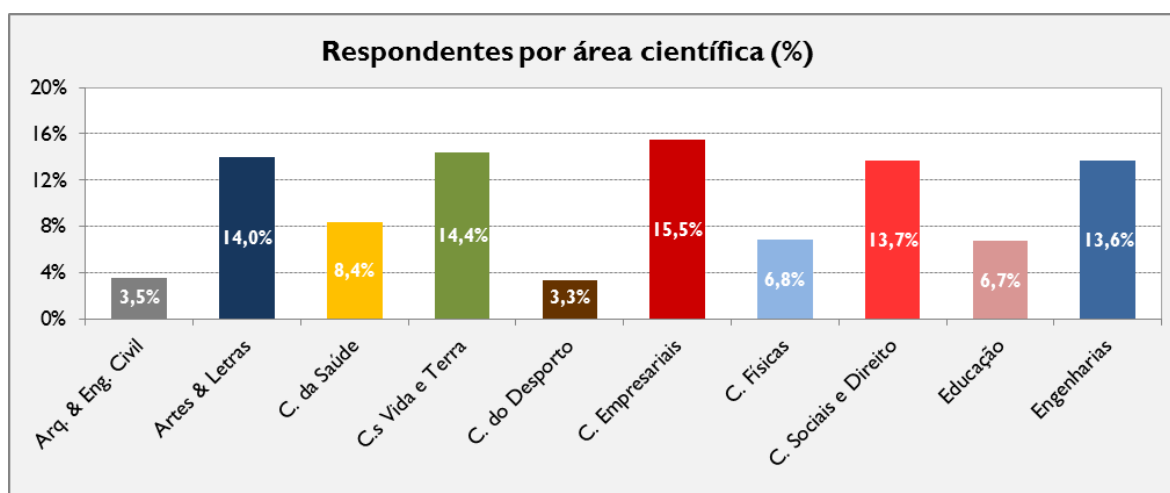


Gráfico 4.4 Distribuição dos respondentes por área científica do curso.

No que respeita à situação profissional no momento do inquérito, através da Tabela 1 é possível observar que 1412 dos 2882 respondentes (cerca de 49%) estava numa situação de desemprego. Na análise da situação profissional por ano letivo, verifica-se que o número de desempregados só é superior ao número de empregados nos diplomados de 2010/2011. O teste de homogeneidade de qui-quadrado (nível de significância de 5%) revela que não existe homogeneidade na distribuição da situação profissional nos três anos letivos ($ET\chi^2 = 144,392$; $p^5 < 0.001$). O coeficiente V de Cramér, cujo valor é de 0.227 ($p < 0.001$), revela que existe alguma relação ainda que reduzida.

⁴ A categoria “Engenharias” não abarca a Engenharia Civil, que está incluída na “Arquitetura & Construção”, mais à frente no estudo da empregabilidade encontram-se definidas cada uma das 10 áreas científicas.

⁵ $ET\chi^2 = 144,392$; $p = 4,424E-32$

A análise dos resíduos estandardizados corroboram estes resultados, assim como permite evidenciar que o número de desempregados diplomados em 2010/2011 ficou acima do esperado. Por sua vez, o número de desempregados entre diplomados de 2008/2009 e de 2009/2010 foi menor do que o valor esperado.

| Situação Profissional | Frequências | Ano letivo | | | Total |
|-----------------------|-------------------------|------------|-----------|-----------|-------|
| | | 2008/2009 | 2009/2010 | 2010/2011 | |
| Desempregado | Observadas | 205 | 397 | 810 | 1412 |
| | Esperadas | 305 | 446 | 662 | 1412 |
| | Resíduos Ajust. Estand. | -9,20 | -4,00 | 11,30 | - |
| Empregado | Observadas | 397 | 484 | 497 | 1378 |
| | Esperadas | 297 | 435 | 646 | 1378 |
| | Resíduos Ajust. Estand. | 9,20 | 4,00 | -11,30 | - |
| Total | Observadas | 602 | 881 | 1307 | 2790 |
| | Esperadas | 602 | 881 | 1307 | 2790 |

Tabela 4.1 Situação profissional dos diplomados (empregado e desempregados, sem os 92 estudantes): frequências observadas e esperadas e resíduos ajustados estandardizados, por ciclos de estudo.

4.2.2. Caracterização por ciclo de estudos

Diplomados de Licenciaturas

No que diz respeito aos 1571 respondentes de 1º Ciclo, são maioritariamente do sexo feminino (59.1%), têm uma média etária de 26.7 anos ($s=5.2$), provêm maioritariamente da Região Centro (76.8%) e apenas cerca de 1% são de países estrangeiros. A sua classificação média final é de 13.2 valores ($s=1.4$).

O Gráfico 4.5 mostra como se distribuem os 1571 licenciados respondentes ao longo dos três anos letivos. A maioria (49.1%) dos respondentes de 1º ciclo são diplomados de 2010/2011, 31.4% são de 2009/2010 e apenas 19.5% são diplomados de 2008/2009.

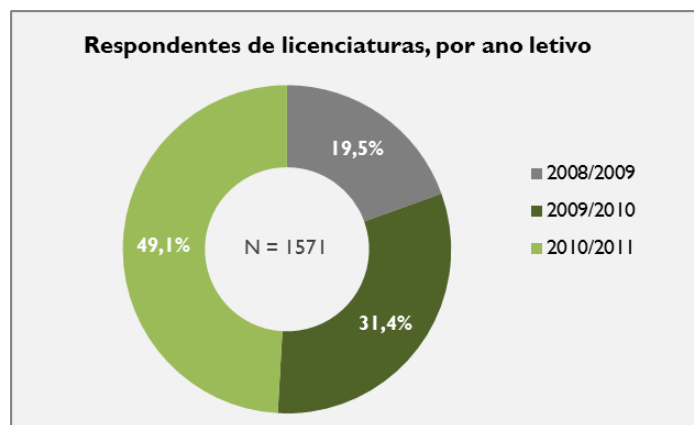


Gráfico 4.5 Distribuição dos respondentes de licenciaturas por ano letivo.

A análise da situação profissional à data do inquérito leva a concluir que entre os 1571 diplomados de licenciatura, 63.7% estavam numa situação de desemprego, 34% referiram ter um emprego regular e 2.4% estudavam. O Gráfico 4.6 permite constatar que quanto mais recente for a graduação, maior a proporção de desempregados licenciados e menor a proporção de empregados.

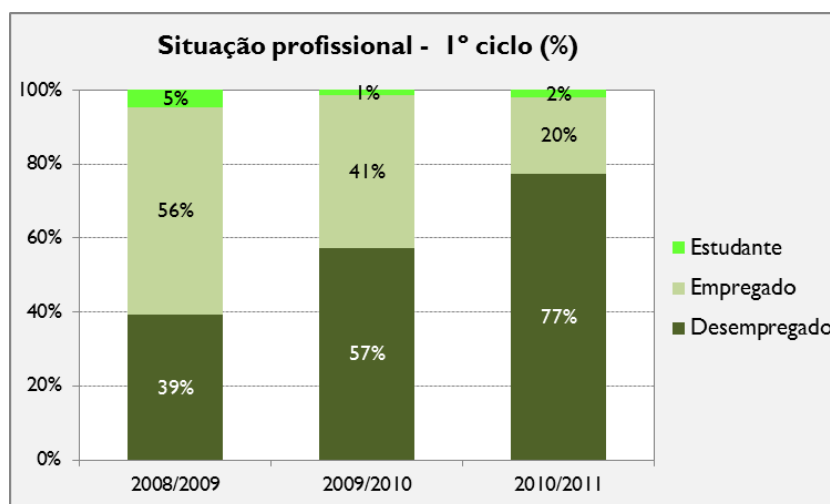


Gráfico 4.6 Distribuição dos respondentes de licenciaturas de acordo com a situação profissional e por ano letivo.

Diplomados de Mestrados Integrados

Entre os 610 respondentes de cursos de 2º ciclo integrado que concluíram entre 2009 e 2011, 52.1% são mulheres, valor inferior relativamente às licenciaturas e também aos mestrados (como se verificará a seguir). A média de idades ronda os 28 anos ($s=3.9$), são predominantemente portugueses (98.7%) e, em grande parte, da Região Centro (cerca de

79.2%). A classificação média de conclusão dos respondentes de Mestrados Integrados é de 14.1 valores ($s=1.4$), sendo assim superior à verificada entre os diplomados de licenciaturas.

A distribuição dos respondentes de mestrados integrados por ano letivo de conclusão está patente no Gráfico 4.7. Embora se verifiquem diferenças nos valores percentuais obtidos nos diferentes anos letivos, esta é inferior à verificada nos diplomados de licenciaturas.

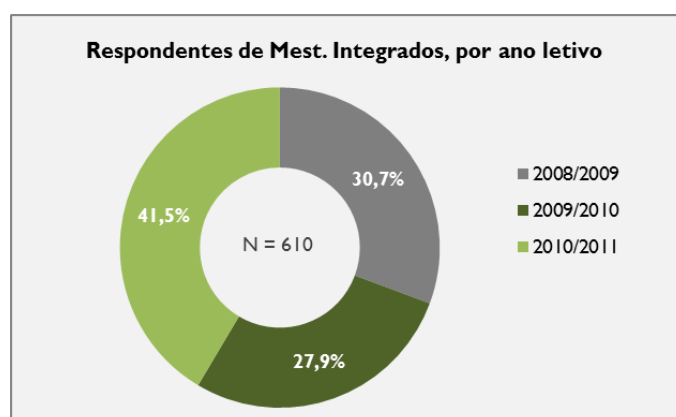


Gráfico 4.7 Distribuição dos respondentes de mestrados integrados por ano letivo.

No que se refere à situação profissional no momento da inquirição, 29% dos 610 diplomados de mestrados integrado estavam numa situação de desemprego, 66.6% tinham um emprego regular e cerca de 4% estudavam. A análise desta questão por ano de conclusão dos diplomados de mestrados integrados é diferente daquela que se verificou entre os licenciados (Gráfico 4.8), há um claro predomínio da situação de emprego nos três anos letivos, embora se continue a observar que a situação de desemprego aumenta entre diplomados mais recentes.

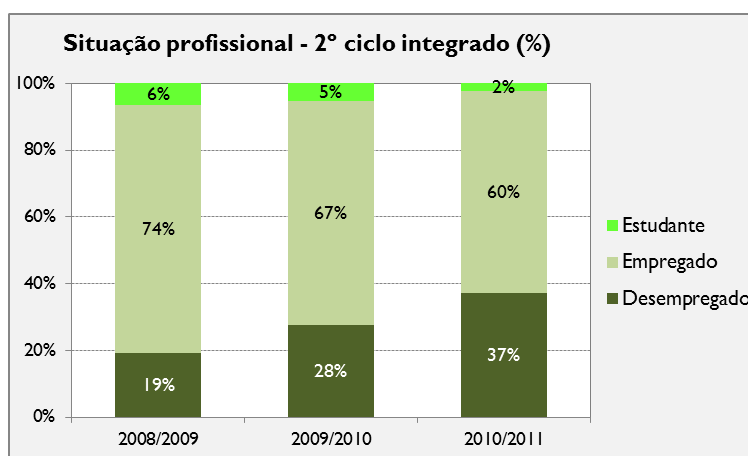


Gráfico 4.8 Distribuição dos respondentes de mestrados integrados de acordo com a situação profissional e por ano letivo.

Diplomados de Mestrados

No que toca aos 701 diplomados de mestrados são, tal como nos casos anteriores, maioritariamente do sexo feminino (57.6%), naturais da região Centro (79.2%), portugueses (94.4%) e têm uma média de idades a rondar os 28 anos ($s=6.3$). A média da classificação final destes diplomados é maior do que a dos diplomados de licenciatura e de mestrados integrados, neste caso ronda os 16 valores ($s=1.4$).

O Gráfico 4.9 permite perceber como se distribuem os diplomados de mestrados pelos três anos letivos, destacando-se novamente o ano letivo 2010/2011 como o mais representado.

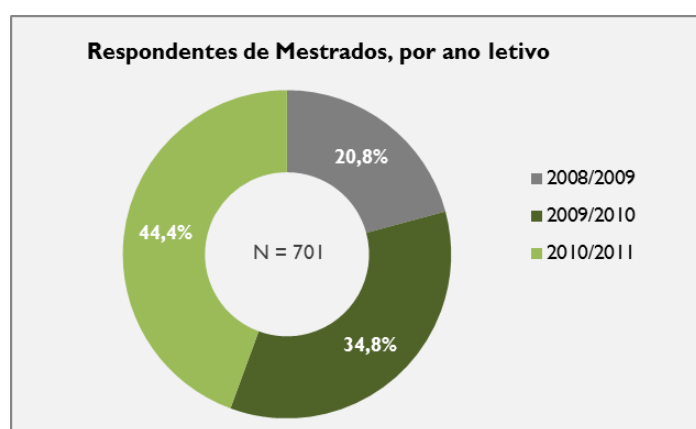


Gráfico 4.9 Distribuição dos respondentes de mestrados por ano letivo.

Relativamente à situação profissional dos respondentes de mestrado, 33.5% estavam, no momento da inquirição, desempregados, 62.5% estavam empregados e 4% continuavam a estudar.

De acordo com o Gráfico 4.10, que reflete a distribuição dos respondentes de mestrados segundo a situação profissional e por ano letivo, verifica-se que, em comparação com os ciclos de estudo anteriores, existe um maior equilíbrio entre as proporções de desempregados e de empregados.

Na comparação com os diplomados de 1º ciclo, registam-se valores inferiores de desemprego entre os diplomados de mestrado, sendo que em 2008/2009 registam-se valores próximos.

Por sua vez, comparando os mestrados integrados e não integrados, constata-se que os valores para a situação de desemprego são muito próximos entre diplomados de 2009/2010 e

de 2010/2011. Já entre os diplomados de 2008/2009 existe uma maior taxa de desemprego nos diplomados de 2º ciclo (34%, em oposição aos 19% do mestrados integrados).

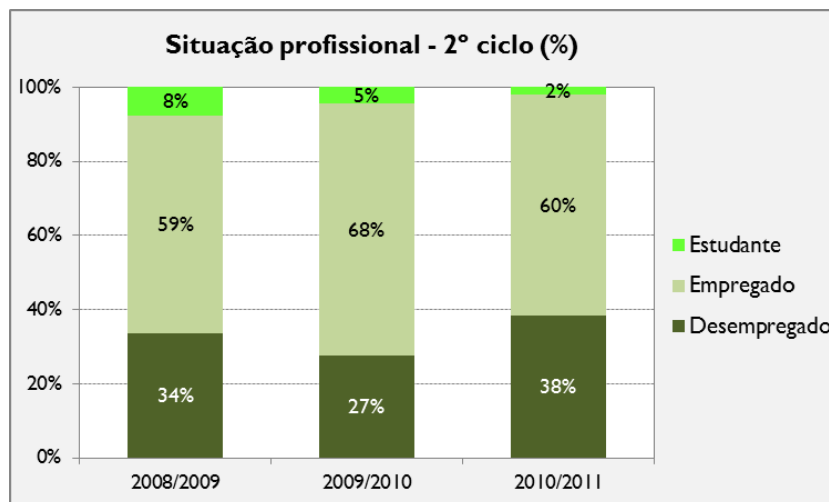


Gráfico 4.10 Distribuição dos respondentes de mestrados de acordo com a situação profissional e por ano letivo.

Em suma, fazendo a comparação dos valores obtidos para as situações profissionais por ciclos de estudo, constata-se que a taxa de emprego é superior entre os diplomados de 2º ciclo, com maior destaque nos mestrados integrados. Verifica-se ainda que a taxa de desemprego entre licenciados é superior à taxa de emprego, ao contrário das de 2º ciclo (integrado ou não). Por cada 100 diplomados de licenciaturas, 64 estão desempregados, por sua vez, por cada 100 diplomados de mestrados integrados ou de mestrados, esse número reduz-se a 29 e 34, respetivamente.

4.3. Percurso Académico

No estudo sobre o emprego/desemprego dos diplomados importa conhecer o percurso académico dos mesmos no sentido de perceber se existe alguma associação entre estas duas questões. Assim, neste ponto analisar-se-á a participação em programas de intercâmbio e em estágios não curriculares, a situação ocupacional no último ano de estudos e se os diplomados prosseguiram mais estudos após a conclusão do curso.

No conjunto dos 2882 respondentes, diplomados entre 2009 e 2011, cerca de 73% declarou que o curso concluído foi a primeira experiência no ensino superior. No entanto, quando se analisa esta questão por ciclos de estudo, verifica-se, como seria expectável, que é entre os respondentes de 1º ciclo e de Mestrados Integrados que se registam os maiores valores, 92.7% e 89.2%, respetivamente.

Ao longo do curso é dada a possibilidade dos estudantes participarem em programas de intercâmbio ou mobilidade. Embora esta participação seja referida e valorizada no suplemento ao diploma, constata-se que a adesão dos respondentes diplomados de 2009 a 2012 foi bastante reduzida (12.6%).

De acordo com os resultados da Tabela 4.2 é possível observar que nos três ciclos de estudos predomina a não participação em programas de intercâmbio, com menor evidência entre os mestrados integrados (1º ciclo: 88.1 | 2º ciclo integrado: 82.1% | 2º ciclo: 90.4%). Para perceber se existem diferenças na participação em programas de mobilidade por ciclo de estudos, recorreu-se ao teste de homogeneidade de qui-quadrado (nível de significância de 5%). Este revelou não existir homogeneidade entre os diplomados que participam e aqueles que não participam em programas de intercâmbio ($ET\chi^2 = 21,968$; $p < 0.001$).

A análise dos resíduos estandardizados permite verificar que é entre os diplomados de mestrados integrados que se verifica uma maior diferença entre os valores observados e os valores esperados, neste caso, o número de diplomados que participaram em programas de mobilidade é superior ao esperado.

| Participação em programa de intercâmbio | | Ciclos de estudos | | | Total |
|--|-------------------------|-------------------|----------------------|-----------|-------|
| | | Licenciaturas | Mestrados Integrados | Mestrados | |
| Não participou em programas de intercâmbio | Observadas | 1384 | 501 | 634 | 2519 |
| | Esperadas | 1373 | 533 | 613 | 2519 |
| | Resíduos Ajust. Estand. | 1,20 | -4,40 | 2,80 | - |
| Participou em programas de intercâmbio | Observadas | 187 | 109 | 67 | 363 |
| | Esperadas | 198 | 77 | 88 | 363 |
| | Resíduos Ajust. Estand. | -1,20 | 4,40 | -2,80 | - |
| Total | Observadas | 1571 | 610 | 701 | 2882 |
| | Esperadas | 1571 | 610 | 701 | 2882 |

Tabela 4.2 Situação profissional dos diplomados: Frequências observadas e esperadas e resíduos ajustados estandardizados, por ciclos de estudo.

A análise da mobilidade segundo o ano letivo (Gráfico 4.11) não revela grandes alterações relativamente ao que se constatou. Nos três anos letivos, mais de 85% dos diplomados não participou em programas de mobilidade, registando-se proporções muito próximas.

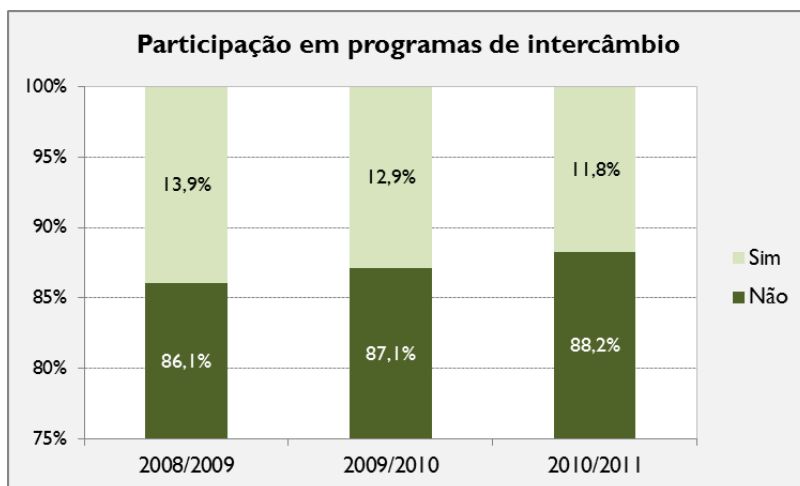


Gráfico 4.11 Participação em programas de intercâmbio por ano letivo.

O teste de homogeneidade não permite concluir de “forma categórica” que a participação ou não em programas de mobilidade se relaciona com a situação profissional dos diplomados ($ET\chi^2 = 4.056$; $p = 0.044$).

Outro aspeto que importa conhecer no estudo da empregabilidade é a situação ocupacional durante o último ano do curso. Para esta questão foram dadas quatro opções de resposta: estudava e realizava trabalho ocasional; estudava e trabalhava a tempo integral; estudava e trabalhava a tempo parcial; e só estudava.

Verifica-se que cerca de 68% dos 2882 diplomados só estudava no seu último ano do curso e 13.3% estudava e acumulava com os estudos um emprego em tempo integral. Os resultados do teste do qui-quadrado revelam que a situação ocupacional no último ano do curso é diferente nos três ciclos de estudo ($ET\chi^2 = 290,315$; $p^6 < 0.001$). A análise do coeficiente V de Cramér corrobora esta conclusão, ainda que a associação seja reduzida ($V = 0.224$; $p = 0.0$).

Relativamente à análise das frequências observadas e esperadas (Tabela 4.3), é possível constatar que a situação ocupacional dos diplomados no último ano de curso está relacionada com o respetivo ciclo de estudos, destacando-se a categoria dos mestrados com o maior

⁶ $ET\chi^2 = 290,315$; $p = 9,716E-60$

contributo nas diferenças encontradas. De facto, o número de diplomados de mestrado que acumulavam os estudos com um emprego a tempo integral supera o valor esperado em mais do dobro. Situação inversa acontece entre os diplomados de licenciatura e de mestrados integrados. Nestes casos, é possível observar que registam valores superiores aos valores esperados na situação em não acumulavam qualquer trabalho com os estudos no último ano do curso.

| Situação ocupacional no último curso | | Ciclos de estudos | | | Total |
|--|-------------------------|-------------------|----------------------|-----------|-------|
| | | Licenciaturas | Mestrados Integrados | Mestrados | |
| Estudava e realizava trabalho ocasional | Observadas | 184 | 43 | 57 | 284 |
| | Esperadas | 155 | 60 | 69 | 284 |
| | Resíduos Ajust. Estand. | 3,70 | -2,60 | -1,80 | - |
| Estudava e trabalhava a tempo completo ou integral | Observadas | 128 | 43 | 216 | 387 |
| | Esperadas | 211 | 82 | 94 | 387 |
| | Resíduos Ajust. Estand. | -9,10 | -5,20 | 15,50 | - |
| Estudava e trabalhava a tempo parcial | Observadas | 154 | 30 | 81 | 265 |
| | Esperadas | 145 | 56 | 65 | 265 |
| | Resíduos Ajust. Estand. | 1,20 | -4,10 | 2,50 | - |
| Só estudava | Observadas | 1105 | 494 | 347 | 1946 |
| | Esperadas | 1061 | 412 | 473 | 1946 |
| | Resíduos Ajust. Estand. | 3,50 | 8,00 | -11,70 | - |
| Total | Observadas | 1571 | 610 | 701 | 2882 |
| | Esperadas | 1571 | 610 | 701 | 2882 |

Tabela 4.3 Situação ocupacional no último ano de curso: Frequências observadas e esperadas e resíduos ajustados estandardizados, por ciclos de estudo.

A análise da situação ocupacional no último ano do curso por ano letivo, patente no Gráfico 4.12, revela um comportamento idêntico nos três anos letivos, salientando-se, entre os diplomados de 2010/2011, uma ligeira diminuição nos valores relativos às situações de acumulação de um trabalho, parcial ou integral, com os estudos.

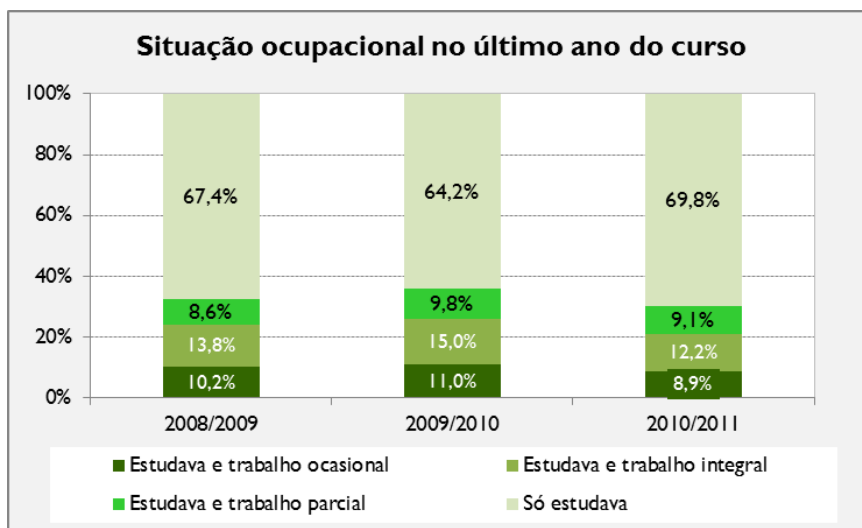


Gráfico 4.12 Distribuição dos respondentes de mestrados de acordo com a situação profissional e por ano letivo.

Na comparação desta variável com a situação profissional dos diplomados o teste de homogeneidade permite concluir que de facto se rejeita a hipótese nula ($ET\chi^2 = 224.714$; $p < 0.001$). O coeficiente V de Cramér, cujo valor é 0.284 ($p < 0.001$) revela que existe alguma relação entre a situação ocupacional dos diplomados no último ano de curso e a situação profissional, favorável aos diplomados mais ativos, ou seja, que acumularam estudos com um trabalho.

Após a conclusão do curso cerca de 71% dos diplomados, 1º ou 2º ciclo, prosseguiram mais estudos no ensino superior. Destes, perto de 68% inscreveram-se num mestrado e neste caso, como se esperaria, destacam-se os diplomados de 1º ciclo (91.8%). Por sua vez, entre os 19% que se inscreveram num curso de doutoramento, há um maior número de diplomados de 2º ciclo.

De acordo com os dados da Tabela 4.4 é possível verificar que, na relação entre estas duas variáveis, há um maior contributo dos diplomados de 2º ciclo (integrado ou não) na não prossecução de estudos. Por outro lado, os diplomados de licenciaturas são os que mais contribuem entre os que prosseguem estudos.

| Progressão de estudos após concluir o curso | | Ciclos de estudos | | | Total |
|---|-------------------------|-------------------|----------------------|-----------|-------|
| | | Licenciaturas | Mestrados Integrados | Mestrados | |
| Não prosseguem estudos | Observadas | 139 | 319 | 381 | 839 |
| | Esperadas | 457 | 204 | 178 | 839 |
| | Resíduos Ajust. Estand. | -26,20 | 11,00 | 20,40 | - |
| Prosseguem estudos | Observadas | 1432 | 382 | 229 | 2043 |
| | Esperadas | 1114 | 497 | 432 | 2043 |
| | Resíduos Ajust. Estand. | 26,20 | -11,00 | -20,40 | - |
| Total | Observadas | 1571 | 701 | 610 | 2882 |
| | Esperadas | 1571 | 701 | 610 | 2882 |

Tabela 4.4 Continuação de estudos no ensino superior, por ciclo.

Recorrendo a uma escala de Likert de 5 pontos (1: Discordo Completamente; 2: Discordo; 3: Nem concordo nem discordo; 4: Concordo; 5: Concordo completamente), tentou perceber-se quais os motivos que levam os diplomados a prosseguir ou não os estudos no ensino superior.

Assim, no que toca aos 2043 diplomados (1º e 2º ciclos) que continuaram os estudos no ensino superior, os fatores que parecem influenciar a sua opção são o aprofundar conhecimentos e competências, o desenvolver competências pessoais e o aumento de oportunidades de emprego. A maioria das respostas concentra-se no “concordo” ou no “concordo completamente”. O item com menor expressão de concordância é o de continuar a ter benefícios como estudante. A análise por ciclos confirma estas constatações anteriores (Tabela 4.5).

| Factores que motivaram a continuação dos estudos | Licenciaturas (N=1432) | | Mestrados Integrados (N=229) | | Mestrados (N=382) | | Total (N=2043) | |
|--|---------------------------|-----|---------------------------------|-----|----------------------|-----|-------------------|-----|
| | Média | DP | Média | DP | Média | DP | Média | DP |
| Para aprofundar conhecimentos e competências | 4,6 | 0,6 | 4,7 | 0,6 | 4,7 | 0,5 | 4,6 | 0,6 |
| Para desenvolver competências pessoais (e.g., empreendedor, comunicação, trabalho de equipa, resolução de problemas) | 4,1 | 0,9 | 4,2 | 0,9 | 4,2 | 0,8 | 4,2 | 0,9 |
| Por considerar insuficiente a formação científica anterior | 3,9 | 1,1 | 3,1 | 1,2 | 2,9 | 1,1 | 3,6 | 1,2 |
| Por influência de outros (família, colegas ou professores) | 2,5 | 1,1 | 2,5 | 1,1 | 2,6 | 1,1 | 2,5 | 1,1 |
| Para continuar a ter benefícios como estudante | 1,9 | 1,0 | 1,7 | 0,9 | 1,8 | 1,0 | 1,8 | 0,9 |
| Porque não tinha emprego | 2,5 | 1,3 | 2,3 | 1,3 | 2,2 | 1,2 | 2,4 | 1,3 |
| Para aumentar as oportunidades de emprego | 4,5 | 0,8 | 4,1 | 1,0 | 4,0 | 1,0 | 4,3 | 0,9 |

Tabela 4.5 Fatores que influenciaram a continuação dos estudos no ensino superior, por ciclo de estudos.

No que concerne aos motivos que levaram os 839 respondentes (1º e 2º ciclos) a não prosseguirem estudos (Tabela 4.6) não existe um fator que se destaque como o motivador da não progressão de estudos. A influência de outros (família, colegas ou professores) não é relevante na tomada de decisão.

Na análise por ciclo dos estudos, é possível observar que a influência de outros continua a não pesar muito na decisão de não prosseguir estudos, quer nos diplomados de 1º ciclo quer nos diplomados de 2º ciclo. No entanto, verificam-se algumas diferenças nos valores médios dos fatores que parecem importar nesta tomada de decisão. A falta de recursos assume uma maior importância nas respostas dos diplomados de 1º ciclo. No caso dos diplomados de 2º ciclo, os fatores que se destacam é o de ter emprego em vista.

| Motivos para não continuar estudos | Licenciaturas (N=139) | | Mestrados Integrados (N=381) | | Mestrados (N=319) | | Total (N=8393) | |
|---|--------------------------|-----|---------------------------------|-----|----------------------|-----|-------------------|-----|
| | Média | DP | Média | DP | Média | DP | Média | DP |
| Falta de recursos económicos | 3,5 | 1,4 | 2,8 | 1,4 | 2,7 | 1,3 | 2,9 | 1,4 |
| Ter um emprego em vista | 2,8 | 1,3 | 3,1 | 1,4 | 3,5 | 1,3 | 3,2 | 1,4 |
| Considerar suficiente a formação dada pelo curso | 2,5 | 1,1 | 3,2 | 1,1 | 3,2 | 1,1 | 3,1 | 1,1 |
| Não pretende continuar a estudar | 2,2 | 1,3 | 2,4 | 1,3 | 2,5 | 1,2 | 2,4 | 1,2 |
| Por influência de outros (família colegas ou professores) | 1,9 | 1,0 | 1,7 | 0,9 | 1,8 | 0,9 | 1,8 | 0,9 |

Tabela 4.6 Motivos para não prosseguir os estudos no ensino superior, por ciclo de estudos.

4.4. Percurso Profissional

4.4.1. Primeiro Emprego

Após a conclusão do curso, 29.2% dos diplomados consegue obter um emprego após a conclusão do curso, 45% ainda procura emprego, 14.1% manteve o emprego que tinha enquanto estudava e 11.7% consegue o seu emprego no local onde estagiou (Gráfico 4.13).

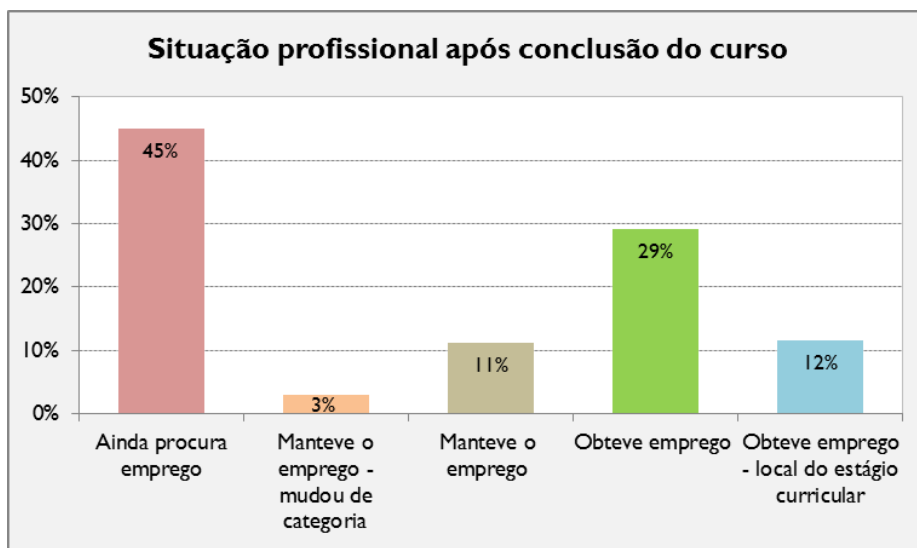


Gráfico 4.13 Situação profissional após a conclusão do curso (N=2882).

Os diplomados que conseguiram efetivamente emprego (retirando aqueles que mantiveram o emprego que tinham enquanto estudantes e os que permaneceram no local onde estagiaram) demoraram em média 5 meses ($s=5.1$) a obtê-lo.

Os licenciados são os que demoraram mais tempo a empregarem-se após conclusão do curso, em média 5.3 meses ($s=5.7$). Seguindo-se os diplomados de mestrados integrados, cujo tempo médio na obtenção do primeiro emprego foi 5.2 meses ($s=5.4$). Os diplomados de cursos de 2º ciclo, refiram que em média demoraram 4.7 meses ($s=4.4$) a conseguir emprego após a conclusão do curso.

A Tabela 4.7 revela que o número de diplomados de licenciaturas desempregados é muito superior ao esperado. No caso dos diplomados de mestrados integrados, observam-se valores acima do esperado nas situações de emprego regular e de emprego decorrente de um estágio. Por sua vez, no caso dos diplomados de mestrados, destacam-se os que mantêm emprego. Recorde-se que este grupo de respondentes também se evidenciou na situação ocupacional do último ano do curso como os que acumulavam os estudos com um trabalho a tempo integral.

| Situação profissional logo após conclusão do curso | | Ciclos de estudos | | | Total |
|--|-------------------------|-------------------|----------------------|--------------|-------|
| | | Licenciaturas | Mestrados Integrados | Mestrados | |
| Ainda à procura de emprego | Observadas | 960 | 142 | 194 | 1296 |
| | Esperadas | 707 | 274 | 315 | 1296 |
| | Resíduos Ajust. Estand. | 19,10 | -12,10 | -10,60 | - |
| Obteve emprego no local de estágio curricular | Observadas | 128 | 107 | 101 | 336 |
| | Esperadas | 183 | 71 | 82 | 336 |
| | Resíduos Ajust. Estand. | -6,40 | 5,10 | 2,60 | - |
| Manteve emprego que tinha enquanto estudava | Observadas | 161 | 42 | 205 | 408 |
| | Esperadas | 222 | 86 | 99 | 408 |
| | Resíduos Ajust. Estand. | -6,60 | -5,80 | 13,20 | - |
| Obteve emprego regular | Observadas | 322 | 319 | 201 | 842 |
| | Esperadas | 459 | 178 | 205 | 842 |
| | Resíduos Ajust. Estand. | -11,30 | 14,10 | -0,40 | - |
| Total | Observadas | 1571 | 610 | 701 | 2882 |
| | Esperadas | 1571 | 610 | 701 | 2882 |

Tabela 4.7 Situação profissional após conclusão do curso, por ciclo.

Como se observa na Tabela 4.8, a grande maioria dos diplomados que consegue ou mantem emprego após conclusão do curso, exerce a profissão na área científica da sua formação (83.6%). No entanto, o comportamento desta variável não é idêntico nos três ciclos de estudo (teste do qui-quadrado: $ET\chi^2 = 67,691$; $p^7 < 0.001$ | coeficiente V de Cramér=0.146; $p=0.000$).

Os diplomados de licenciatura empregados em profissões de áreas distintas da sua formação, e que não requerem formação superior, registam valores superiores ao esperado. Por sua vez, os diplomados de mestrados (integrados ou não) empregados em profissões da sua área de formação obtiveram valores superiores ao esperado.

⁷ $ET\chi^2 = 67,691$; $p=6,969E-14$

| 1º emprego após conclusão do curso na área científica do curso? | | Ciclos de estudos | | | Total |
|---|-------------------------|-------------------|----------------------|-------------|-------|
| | | Licenciaturas | Mestrados Integrados | Mestrados | |
| Não e não requer formação superior | Observadas | 97 | 17 | 25 | 139 |
| | Esperadas | 54 | 41 | 44 | 139 |
| | Resíduos Ajust. Estand. | 7,90 | -4,70 | -3,70 | - |
| Não, mas requer formação superior | Observadas | 53 | 30 | 38 | 121 |
| | Esperadas | 47 | 36 | 39 | 121 |
| | Resíduos Ajust. Estand. | 1,20 | -1,20 | -0,10 | - |
| Sim | Observadas | 461 | 421 | 444 | 1326 |
| | Esperadas | 511 | 391 | 424 | 1326 |
| | Resíduos Ajust. Estand. | -6,90 | 4,40 | 2,90 | - |
| Total | Observadas | 611 | 468 | 507 | 1586 |
| | Esperadas | 611 | 468 | 507 | 1586 |

Tabela 4.8 Área do 1º emprego após a conclusão do curso coincidente ou não com área de formação, por ciclo de estudos.

4.4.2. Diplomados empregados

À data do inquérito, 1378 dos 2882 diplomados (47.8%) estava empregado. Destes 51.7% são mulheres, a média etária ronda os 29 anos ($s=5.9$), apenas 27.9% não concluiu o curso no tempo previsto e a média de classificação final foi de 14.4 valores ($s=1.9$).

No que concerne à condição laboral destes diplomados (Gráfico 4.14), é possível verificar um predomínio de trabalhadores por conta de outrem (56%), seguindo-se os bolseiros de investigação (21.4%). Cerca 1% dos diplomados referiram estar num estágio profissional não remunerado.

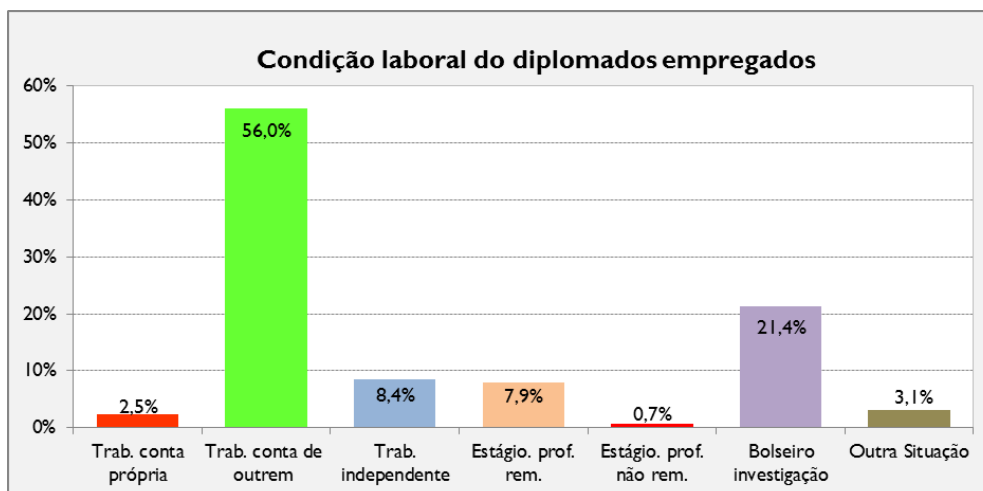


Gráfico 4.14 Condição laboral dos diplomados empregados (N=1378).

Da análise por ciclos de estudos, refletida na Tabela 4.9, constata-se que se registam valores idênticos nos três ciclos para “trabalhadores por conta de outrem” e “trabalhadores independentes”. Por sua vez, os “estágios profissionais remunerados” assumem uma maior proporção entre os diplomados de licenciaturas e no caso dos diplomados de mestrados, integrados ou não, destacam-se os “bolseiro de investigação”.

| Condição Laboral | | Ciclos de estudos | | |
|--------------------------------------|---|-------------------|----------------------|-----------|
| | | Licenciaturas | Mestrados Integrados | Mestrados |
| Trabalhador por conta própria | N | 16 | 7 | 11 |
| | % | 3,0% | 1,7% | 2,5% |
| Trabalhador por conta de outrem | N | 295 | 239 | 237 |
| | % | 55,2% | 58,9% | 54,1% |
| Trabalhador independente | N | 47 | 34 | 35 |
| | % | 8,8% | 8,4% | 8,0% |
| Estágio profissional remunerado | N | 66 | 36 | 7 |
| | % | 12,4% | 8,9% | 1,6% |
| Estágio profissional não remunerado | N | 6 | 3 | 1 |
| | % | 1,1% | 0,7% | 0,2% |
| Bolsheiro de investigação científica | N | 82 | 83 | 130 |
| | % | 15,4% | 20,4% | 29,7% |
| Outra Situação | N | 22 | 4 | 17 |
| | % | 4,1% | 1,0% | 3,9% |
| Total | N | 534 | 406 | 438 |
| | % | 100% | 100,0% | 100,0% |

Tabela 4.9 Área Condição laboral dos diplomados empregados, por ciclos de estudos.

Atendendo ao Gráfico 4.15, é possível concluir que o comportamento desta variável pouco se altera ao longo dos três anos letivos, registando-se apenas ligeiras diferenças nos seus valores.

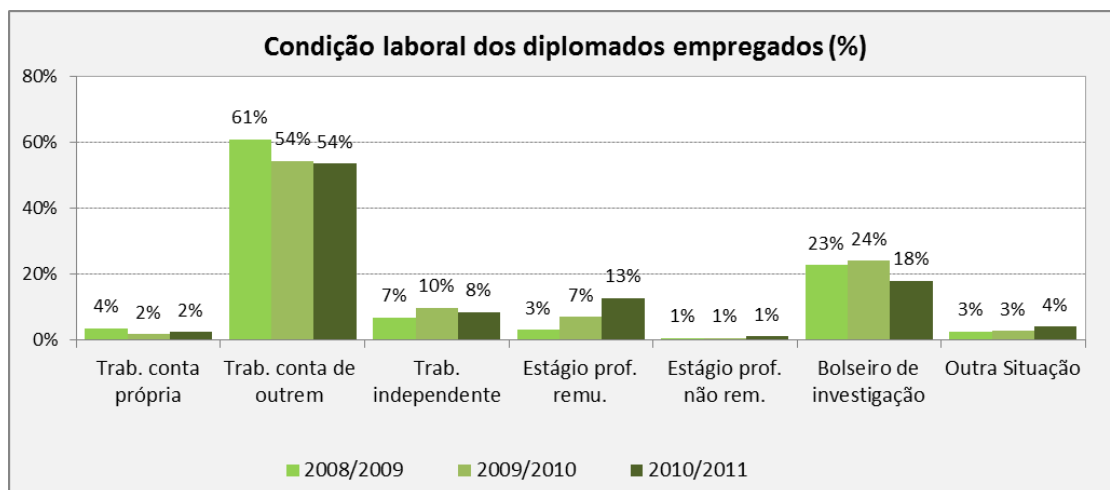


Gráfico 4.15 Condição laboral dos diplomados empregados, por ano letivo.

Nas situações de emprego entre os diplomados importa averiguar qual a relação entre o grau dos diplomados (ciclo de estudo) e o respetivo rendimento auferido. Assim, de acordo com a Tabela 4.10, é possível concluir que os licenciados são os que mais contribuem para as remunerações mais baixas. Já os mestres assumem um maior contributo nas remunerações mais elevadas, onde se destacam os diplomados de mestrados integrados nas remunerações acima dos 1400€.

| Rendimento líquido mensal (€) | | Ciclos de estudos | | | Total |
|-------------------------------|-------------------------|-------------------|----------------------|-------------|-------|
| | | Licenciaturas | Mestrados Integrados | Mestrados | |
| Não Remunerado | Observadas | 6 | 3 | 1 | 10 |
| | Esperadas | 4 | 3 | 3 | 10 |
| | Resíduos Ajust. Estand. | 1,40 | 0,00 | -1,50 | - |
| ≤ 500 | Observadas | 70 | 15 | 21 | 106 |
| | Esperadas | 41 | 31 | 34 | 106 |
| | Resíduos Ajust. Estand. | 6,00 | -3,60 | -2,80 | - |
| 501 – 800 | Observadas | 161 | 87 | 75 | 323 |
| | Esperadas | 125 | 95 | 103 | 323 |
| | Resíduos Ajust. Estand. | 4,70 | -1,10 | -3,80 | - |
| 801 – 1100 | Observadas | 171 | 151 | 199 | 521 |
| | Esperadas | 202 | 154 | 166 | 521 |
| | Resíduos Ajust. Estand. | -3,50 | -0,30 | 4,00 | - |
| 1101 – 1400 | Observadas | 55 | 91 | 60 | 206 |
| | Esperadas | 80 | 61 | 66 | 206 |
| | Resíduos Ajust. Estand. | -3,90 | 5,00 | -0,90 | - |
| 1401 – 1700 | Observadas | 22 | 15 | 30 | 67 |
| | Esperadas | 26 | 20 | 21 | 67 |
| | Resíduos Ajust. Estand. | -1,00 | -1,30 | 2,30 | - |
| ≥ 1701 | Observadas | 23 | 29 | 39 | 91 |
| | Esperadas | 35 | 27 | 29 | 91 |
| | Resíduos Ajust. Estand. | -2,70 | 0,50 | 2,30 | - |
| N/R | Observadas | 26 | 15 | 13 | 54 |
| | Esperadas | 21 | 16 | 17 | 54 |
| | Resíduos Ajust. Estand. | 1,40 | -0,30 | -1,20 | - |
| Total | Observadas | 534 | 406 | 438 | 1378 |
| | Esperadas | 534 | 406 | 438 | 1378 |

Tabela 4.10 Área Condição laboral dos diplomados empregados, por ciclos de estudos.

É usual pensar que existem áreas científicas com maior tendência para empregar diplomados do que outras. Consequentemente, para se analisar esta questão entre os diplomados da instituição de ensino superior, os cursos foram categorizados em 10 áreas dominantes definidas da seguinte forma:

- Arquitetura e Construção: compreende cursos relacionados com a arquitetura e engenharia civil;
- Artes e Humanidades: abrange os cursos das letras, artes e humanidades;
- Ciência da Saúde: inclui os cursos da área da medicina e farmácia;
- Ciências da Vida e da Terra: agrupa os cursos ligados às ciências da natureza;

- Ciências do Desporto: categoriza os cursos respeitantes à área do desporto;
- Ciências Empresariais: abarca os cursos relacionados com o mundo empresarial (economia, gestão, relações internacionais, etc.) assim como a licenciatura de administração público-privada;
- Ciências Físicas: reúne os cursos relacionados com a matemática, física e química;
- Ciências Sociais e Direito: neste grupo incluem-se os cursos de cariz social e humano como a sociologia ou a psicologia, assim como o direito;
- Educação: agrupa todos os cursos de ensino bem como as ciências da educação;
- Engenharias e Tecnologia: inclui todos os cursos de engenharias (excetuando a engenharia civil) e os relacionados com as tecnologias da informação.

O Gráfico 4.16 espelha a empregabilidade dos diplomados de 2009 a 2011 da instituição segundo a área científica do curso. Assim, destacam-se pela positiva os cursos de Engenharias e da área das Ciências Empresariais que representam, respetivamente, 21% e 15% da empregabilidade dos diplomados em análise. Em sentido oposto, destacam-se as Ciências do Desporto, com 3%, e a Arquitetura & Construção, cujo valor é de 4%.

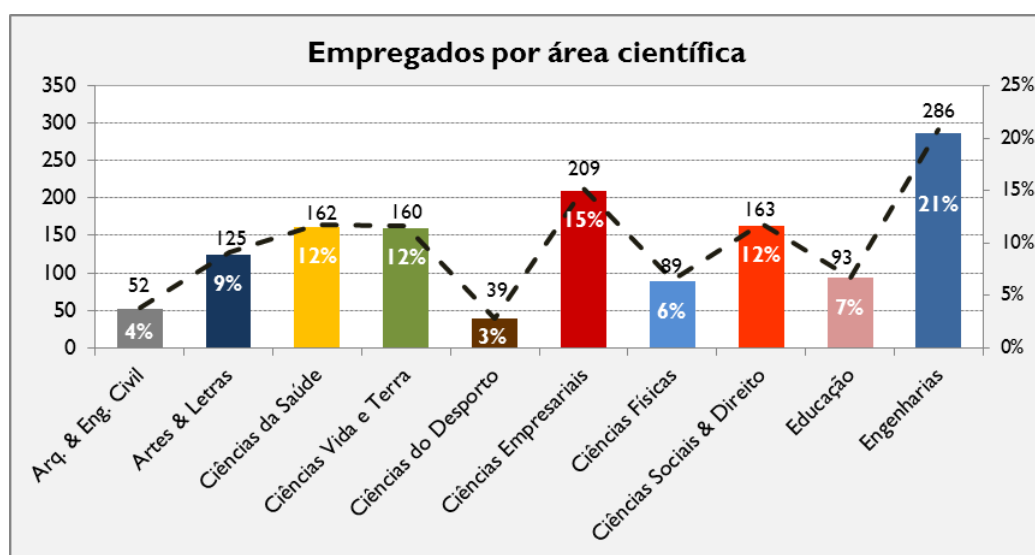


Gráfico 4.16 Diplomados empregados segundo a área científica do curso (N=1378).

A análise segundo o ano letivo não revela grandes diferenças ao que se conclui anteriormente. De facto o Gráfico 4.17 mostra que a distribuição dos diplomados empregados por área

científica nos diferentes três anos letivos segue o mesmo comportamento, ainda com ligeiras diferenças percentuais.

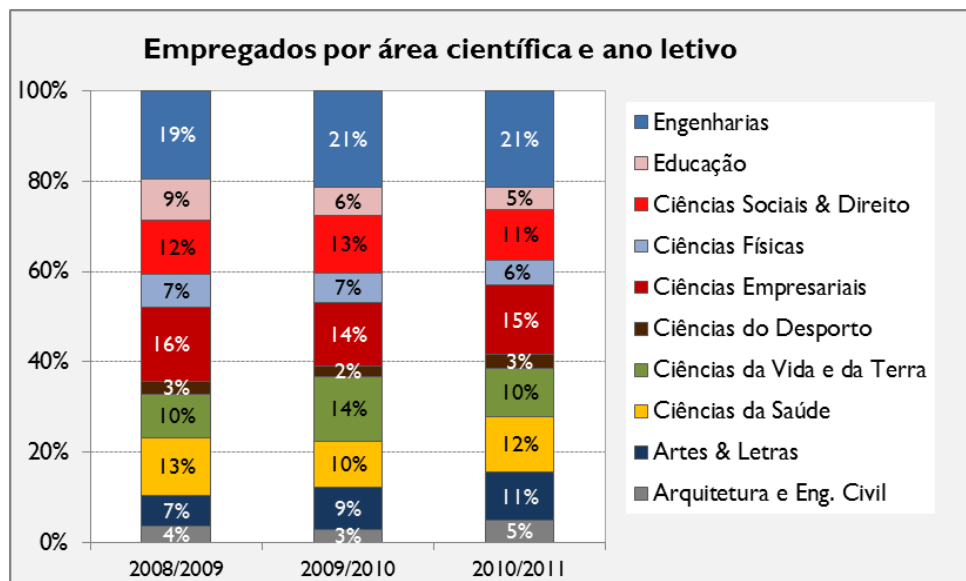


Gráfico 4.17 Diplomados empregados segundo a área científica do curso, por ano letivo (N=1378).

4.4.3. Diplomados desempregados

No momento da inquirição cerca de 49% dos 2882 respondentes (1412) referiu estar numa situação de desemprego. Verifica-se que a maioria é do género feminino⁸ (62,7%) superando a proporção verificada entre os respondentes empregados (51.7%). A média de idades ronda os 27 anos ($s=4.8$), ou seja, comparativamente com os respondentes empregados (27.9 anos), estes tendem a ser mais jovens. Apenas 23.4% dos respondentes desempregados não concluiu o curso no tempo previsto, sendo a média de classificação final destes diplomados igual a 13.9% ($s=1.8$), ligeiramente mais baixa que a verificada entre os empregados (14.4 valores).

No que respeita à distribuição dos desempregados pelas áreas científica dos cursos (Gráfico 18), verifica-se que as Ciências Sociais e Direito surgem entre as áreas mais representadas (15.9%) só é superada pelas Artes e Humanidades (19.1%). Em sentido oposto destacam-se a Arquitetura e Engenharia Civil e as Ciências do Desporto, que registam valores muito próximos aos observados no estudo da empregabilidade significando que os respondentes destas áreas

⁸O teste de homogeneidade ($ET_{\chi^2}=34.980$; $p<0.001$) corrobora esta constatação. Não existe homogeneidade entre a variável género e a situação profissional, concluindo-se que o desemprego atinge mais as mulheres.

registam uma distribuição idêntica no que diz respeito ao número de empregados e de desempregados.

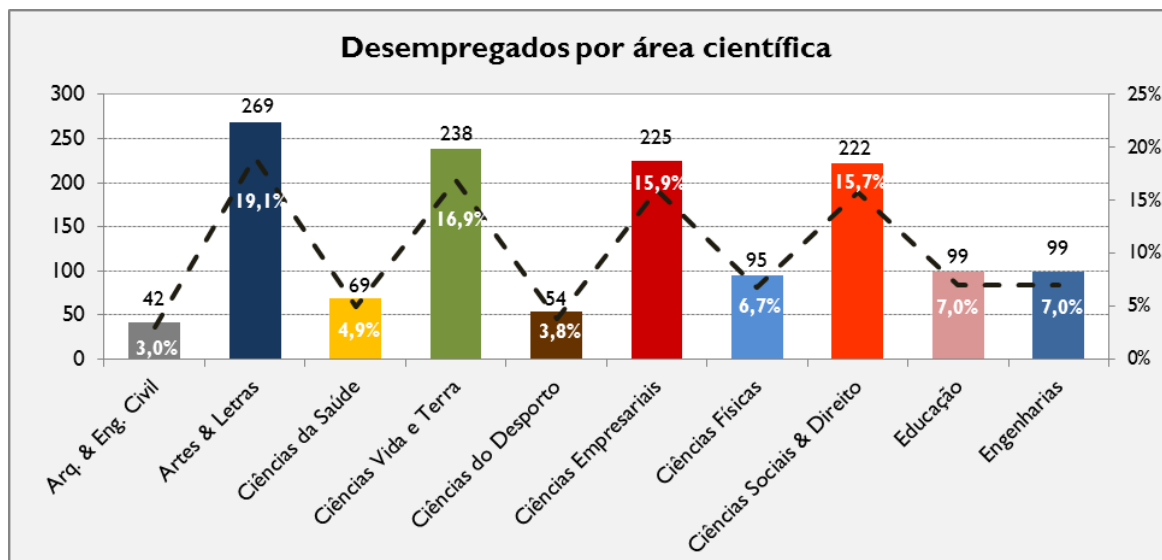


Gráfico 4.18 Diplomados desempregados segundo a área científica do curso (N=1412).

De forma a entender as dificuldades sentidas pelos diplomados desempregados na obtenção de um emprego, considerou-se um conjunto de cinco itens/motivos que foram avaliados de acordo com uma escala de Likert de 5 pontos (1: Discordo Completamente; 2: Discordo; 3: Nem concordo nem discordo; 4: Concordo; 5: Concordo completamente). A Tabela 4.11 espelha os resultados desta análise. Genericamente é possível concluir que as três principais dificuldades são a falta de experiência profissional (com maior destaque entre os licenciados), a falta de oferta de empregos na área científica e o excesso de diplomados na área científica. A indisponibilidade de afastamento da residência e as condições remuneratórias não assumiram particular importância entre os diplomados desempregados.

| Dificuldades na obtenção de emprego | Licenciaturas (N=997; NR=3) | | Mestrados Integrados (N=176; NR=1) | | Mestrados (N=234; NR=1) | | Total (N=1407; NR=5) | |
|--|--------------------------------|-----|---------------------------------------|-----|----------------------------|-----|-------------------------|-----|
| | Média | DP | Média | DP | Média | DP | Média | DP |
| Falta de experiência profissional | 3,7 | 1,1 | 3,4 | 1,3 | 3,0 | 1,3 | 3,5 | 1,2 |
| Fraca oferta de empregos na área científica | 3,8 | 1,1 | 4,0 | 1,0 | 4,0 | 1,0 | 3,9 | 1,1 |
| Excesso de diplomados na área científica | 3,5 | 1,1 | 4,0 | 1,1 | 3,5 | 1,1 | 3,5 | 1,1 |
| Indisponibilidade para me afastar da residência habitual | 2,0 | 1,0 | 1,8 | 0,9 | 1,9 | 1,0 | 2,0 | 1,0 |
| Más condições remuneratórias de outras propostas | 3,0 | 1,0 | 3,2 | 1,2 | 2,8 | 1,1 | 3,0 | 1,1 |

Tabela 4.11 Razões das dificuldades na obtenção de emprego.

A distribuição dos desempregados diplomados difere pouco ao longo dos três anos letivos. Atendendo ao Gráfico 4.19, destaca-se em 2010/2011 o aumento da proporção de desempregados das áreas das Engenharias, Ciências Sociais e Direito e das Ciências Empresarias.

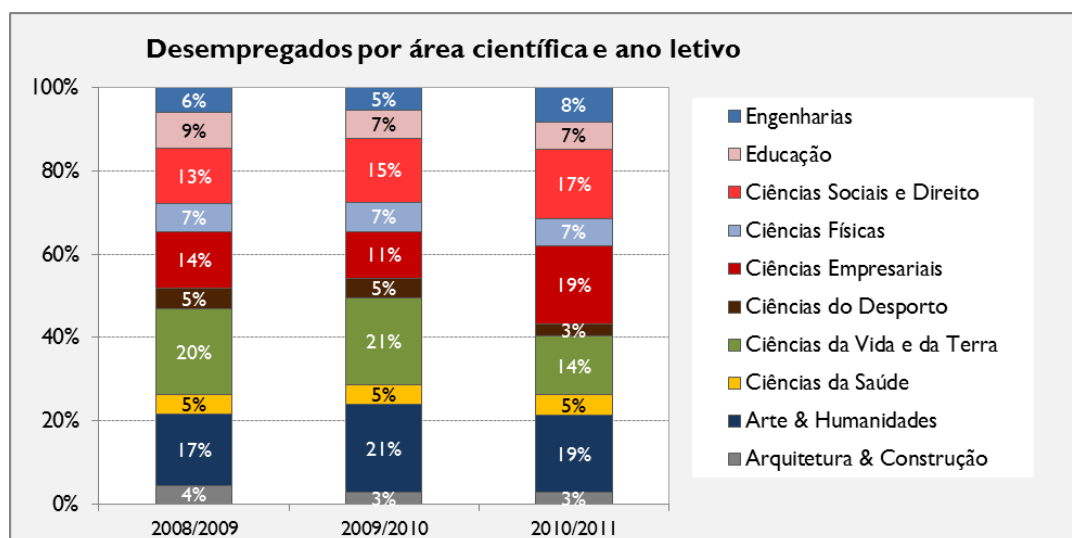


Gráfico 4.19 Diplomados desempregados segundo a área científica do curso e ano letivo.

4.5. Modelo logístico da empregabilidade

Depois de analisar o percurso académico dos diplomados e como este se cruza com a situação profissional, nesta secção pretende-se saber se há condições que favorecem as situações de emprego/desemprego. Assim, considerando algumas das variáveis, sociodemográficas e académicas, tidas normalmente como promotoras de emprego e confirmadas pela análise anterior, avalia-se a sua influência na situação profissional dos diplomados. Entre estas variáveis destacam-se o género, idade, área científica do curso, o ciclo de estudos (licenciatura, mestrado ou mestrado integrado), participação em programas de intercâmbio, participação em estágios não curriculares, a situação ocupacional no último ano de curso e a progressão ou não de estudos após a conclusão do curso.

Com base nos dados recolhidos entre os diplomados⁹, pretende-se construir um modelo que descreva o comportamento da empregabilidade, onde variável de resposta do modelo, Y , assume o valor 1 no caso de diplomado estar empregado (sucesso) e o valor 0, no caso de o diplomado estar desempregado (insucesso). Já as variáveis atrás mencionadas que se julga que poderão estar relacionadas com a situação de emprego/desemprego, constituem as variáveis explicativas do modelo de regressão.

Uma vez que a qualidade do modelo de regressão poderá ser influenciada pela presença de variáveis quantitativas, estas foram transformadas em variáveis qualitativas tal como se apresenta na Tabela 4.12.

⁹ Consideram-se apenas os diplomados empregados ou desempregados (excluem-se os estudantes).

| Variáveis | Categorias |
|--|--|
| X ₁ - Género | 0 - Masculino 1 - Feminino |
| X ₂ - Idade | 0 - Inferior a 30 anos 1 - Superior ou igual a 30 anos |
| X ₃ - Área Científica | 1 - Artes & Humanidades 2 - Ciências da Vida e da Terra 3 - Ciências da Saúde 4 - Ciências do Desporto 5 - Ciências Empresariais 6 - Ciências Físicas 7 - Ciências Sociais e Direito 8 - Educação 9 - Arquitetura e Construção 10 - Engenharias |
| X ₄ - Ciclo de Estudos | 1 - 1º Ciclo de Estudos (Licenciaturas) 2 - 2º Ciclo de Estudos (Mestrados não Integrados) 3 - 2º Ciclo Integrado de Estudos (Mestrados integrados) |
| X ₅ - Participação em programas de mobilidade | 0 - Não 1 - Sim |
| X ₆ - Participação Estágio não curriculares | 0 - Não 1 - Sim |
| X ₇ - Prossegue estudos após conclusão de curso | 0 - Não 1 - Sim |
| X ₈ - Situação ocupacional no último ano do curso | 1 - Estudava e realizava trabalho ocasional 2 - Estudava e trabalhava a tempo integral 3 - Estudava e trabalhava a tempo parcial 4 - Só estudava |

Tabela 4.12 Variáveis relevantes no estudo da empregabilidade.

À partida, na análise dos modelos logísticos, existe sempre uma categoria de referência, que corresponderá àquela que denota a ausência da característica. No caso das variáveis dicotómicas, a categoria de referência é aquela que corresponde à ausência (valor nulo). No caso de existirem mais de duas categorias, de acordo com o que foi abordado no capítulo 4, estas¹⁰ darão azo a novas variáveis, artificiais ou *dummies*, cujos valores são 0 ou 1. O *software* permite obter este procedimento de forma automática, considerando ainda uma das categorias como de referência. Considere-se a variável área do curso como exemplo, esta variável compreende 10 categorias, no entanto, nos resultados da aplicação da regressão

¹⁰ As variáveis ciclo de estudos, área científica do curso e situação ocupacional no último ano do curso.

logística, apenas serão apresentadas 9 áreas, aquela que não consta será a de referência (neste caso será a área das Engenharias).

A procura do modelo logístico inicia-se com a análise individual de cada variável explicativa obtendo-se os resultados dispostos na Tabela 4.13. Esta tabela compreende todas as variáveis candidatas a incluir no modelo e os respetivos parâmetros estimados. Através dos valores de *p-value* dos testes de Wald (correspondentes aos níveis de significância dos coeficientes de regressão) consideram-se as variáveis candidatas a entrar no modelo regressão múltipla com valores inferiores ou iguais a 0.25. Quanto menor for este valor tanto mais importante será considerada a variável. Assim, as variáveis relativas à participação em estágios e à idade parecem não apresentar forte influência na variável resposta.

| Variáveis | B | S.E. | Wald | df | Sig. | Exp(B) | Intervalo Confiança | |
|-------------------------------------|--------|-------|---------|----|--------------|--------|---------------------|-------------|
| | | | | | | | L. Inferior | L. Superior |
| Área Científica | | | 100,983 | 9 | 0,000 | | | |
| Artes e Humanidades | -1,444 | 0,181 | 63,775 | 1 | 0,000 | 0,236 | 0,165 | 0,336 |
| Ciências da Vida e da Terra | -1,012 | 0,176 | 32,928 | 1 | 0,000 | 0,363 | 0,257 | 0,514 |
| Ciências da Saúde | -0,278 | 0,204 | 1,853 | 1 | 0,173 | 0,758 | 0,508 | 1,13 |
| Ciências do Desporto | -1,104 | 0,266 | 17,187 | 1 | 0,000 | 0,331 | 0,197 | 0,559 |
| Ciências Empresariais | -0,628 | 0,172 | 13,283 | 1 | 0,000 | 0,533 | 0,38 | 0,748 |
| Ciências Físicas | -1,275 | 0,261 | 23,77 | 1 | 0,000 | 0,28 | 0,167 | 0,467 |
| Ciências Sociais e Direito | -1,012 | 0,164 | 38,093 | 1 | 0,000 | 0,363 | 0,264 | 0,501 |
| Educação | -0,844 | 0,217 | 15,178 | 1 | 0,000 | 0,43 | 0,281 | 0,657 |
| Arquitetura e Construção | -1,403 | 0,248 | 31,913 | 1 | 0,000 | 0,246 | 0,151 | 0,4 |
| Ciclos de estudo | | | 69,94 | 2 | 0,000 | | | |
| Licenciatura | -1,104 | 0,169 | 42,758 | 1 | 0,000 | 0,332 | 0,238 | 0,462 |
| Mestrado | -0,285 | 0,177 | 2,593 | 1 | 0,107 | 0,752 | 0,532 | 1,064 |
| Género (Feminino) | -0,166 | 0,096 | 2,994 | 1 | 0,084 | 0,847 | 0,703 | 1,022 |
| Idade (≥30 anos) | -0,07 | 0,124 | 0,319 | 1 | 0,572 | 0,932 | 0,731 | 1,189 |
| Estágio não curricular (Sim) | -0,08 | 0,136 | 0,344 | 1 | 0,558 | 0,923 | 0,707 | 1,206 |
| Programa de mobilidade (Sim) | 0,449 | 0,13 | 11,959 | 1 | 0,001 | 1,566 | 1,215 | 2,02 |
| Situação ocupacional | | | 155,212 | 3 | 0,000 | | | |
| Trab. Ocasional | 0,704 | 0,141 | 24,869 | 1 | 0,000 | 2,023 | 1,534 | 2,668 |
| Trab. Integral | 2,042 | 0,172 | 140,484 | 1 | 0,000 | 7,709 | 5,5 | 10,806 |
| Trab. Parcial | 0,778 | 0,151 | 26,73 | 1 | 0,000 | 2,178 | 1,622 | 2,926 |
| Prosegue Estudos (Sim) | -0,346 | 0,11 | 9,847 | 1 | 0,002 | 0,707 | 0,57 | 0,878 |
| Constante | 1,394 | 0,159 | 77,023 | 1 | 0,000 | 4,03 | | |

Tabela 4.13 Aplicação da regressão logística simples.

Excluindo as variáveis acima identificadas e aplicando o método até se conseguir obter um modelo em que o nível de significância para o coeficiente das variáveis explicativas seja significativo ($p < 0.05$) obtém-se o resultado patente na Tabela 4.15, dando origem a um modelo que considera as variáveis área do curso, ciclos de estudo, participação em programa de

intercâmbios, situação ocupacional no último ano e progressão de estudos após a conclusão do curso. Relativamente à aplicação de regressão logística simples (onde se retiraram as variáveis da idade e da participação em estágios não curriculares), foi retirada a variável género por registar um *p-value* superior a 0.05.

O modelo alcançado consegue extrair 70.8% da variabilidade da variável resposta. De acordo com o teste de Hosmer-Lemeshow (Tabela 4.14) verifica-se que os valores estimados não são significativamente diferentes dos observados, o nível de significância igual a 0.504 confirma esta constatação.

| Emprego = 0 | | | Emprego = 1 | | |
|-------------|------------|-----------|-------------|-----------|-------|
| Decil | Observadas | Esperadas | Observadas | Esperadas | Total |
| 1 | 207 | 203,2 | 40 | 43,8 | 247 |
| 2 | 252 | 243,7 | 65 | 73,3 | 317 |
| 3 | 248 | 248,5 | 103 | 102,5 | 351 |
| 4 | 167 | 167,4 | 97 | 96,6 | 264 |
| 5 | 174 | 177,2 | 148 | 144,8 | 322 |
| 6 | 122 | 130,3 | 156 | 147,7 | 278 |
| 7 | 91 | 102,4 | 189 | 177,6 | 280 |
| 8 | 80 | 75,9 | 203 | 207,1 | 283 |
| 9 | 52 | 49,7 | 230 | 232,3 | 282 |
| 10 | 19 | 13,8 | 147 | 152,2 | 166 |

| Qui-quadrado | df | Sig. |
|--------------|----|-------|
| 7,303 | 8 | 0,504 |

Tabela 4.14 Resultados do teste Hosmer-Lemeshow.

| Variáveis | B | S.E. | Wald | df | Sig. | Exp(B) | Intervalo Confiança | |
|-------------------------------------|--------|-------|---------|----|--------------|--------|---------------------|-------------|
| | | | | | | | L. Inferior | L. Superior |
| Área Científica | | | 119,004 | 9 | 0,000 | | | |
| Artes e Humanidades | -1,537 | 0,173 | 78,563 | 1 | 0,000 | 0,215 | 0,153 | 0,302 |
| Ciências da Vida e da Terra | -1,107 | 0,168 | 43,209 | 1 | 0,000 | 0,33 | 0,237 | 0,46 |
| Ciências da Saúde | -0,349 | 0,2 | 3,041 | 1 | 0,081 | 0,705 | 0,477 | 1,044 |
| Ciências do Desporto | -1,126 | 0,267 | 17,841 | 1 | 0,000 | 0,324 | 0,192 | 0,547 |
| Ciências Empresariais | -0,709 | 0,167 | 18,05 | 1 | 0,000 | 0,492 | 0,355 | 0,683 |
| Ciências Físicas | -1,366 | 0,256 | 28,5 | 1 | 0,000 | 0,255 | 0,155 | 0,421 |
| Ciências Sociais e Direito | -1,112 | 0,155 | 51,686 | 1 | 0,000 | 0,329 | 0,243 | 0,445 |
| Educação | -0,959 | 0,207 | 21,555 | 1 | 0,000 | 0,383 | 0,256 | 0,574 |
| Arquitetura e Construção | -1,424 | 0,247 | 33,163 | 1 | 0,000 | 0,241 | 0,148 | 0,391 |
| Ciclos de estudo | | | 70,593 | 2 | 0,000 | | | |
| Licenciatura | -1,071 | 0,167 | 40,975 | 1 | 0,000 | 0,343 | 0,247 | 0,476 |
| Mestrado | -0,254 | 0,174 | 2,131 | 1 | 0,144 | 0,776 | 0,552 | 1,091 |
| Programa de mobilidade (Sim) | 0,439 | 0,129 | 11,624 | 1 | 0,001 | 1,551 | 1,205 | 1,996 |
| Situação ocupacional | | | 184,44 | 3 | 0,000 | | | |
| Trab. Ocasional | 0,706 | 0,14 | 25,416 | 1 | 0,000 | 2,026 | 1,54 | 2,666 |
| Trab. Integral | 2,014 | 0,156 | 165,546 | 1 | 0,000 | 7,49 | 5,512 | 10,179 |
| Trab. Parcial | 0,771 | 0,149 | 26,931 | 1 | 0,000 | 2,162 | 1,616 | 2,894 |
| Prossegue Estudos (Sim) | -0,346 | 0,11 | 9,902 | 1 | 0,002 | 0,707 | 0,57 | 0,878 |
| Constante | 1,325 | 0,154 | 74,011 | 1 | 0,000 | 3,763 | | |

Tabela 4.15 Valores das variáveis explicativas no modelo de regressão logística.

Atendendo a estes resultados, destacam-se as seguintes conclusões:

- Os diplomados que participaram em programas de mobilidade têm mais probabilidade de obter emprego do que aqueles que não participaram (em mais de 50%);
- Os diplomados de licenciatura têm menos de 65% de possibilidade que aqueles que concluíram um curso de mestrado integrado;
- Os diplomados que acumulavam os estudos com um trabalho a tempo integral têm mais hipótese, em mais de 7 vezes, de conseguir emprego que aqueles que só estudam;
- Os diplomados que acumulavam os estudos com um trabalho a tempo parcial ou ocasional têm em mais do dobro de hipóteses de emprego do que aqueles que só estudam;
- Os diplomados que continuam os estudos após a conclusão do curso têm menos de cerca de 29% de possibilidade do que aqueles que não continuam os estudos;
- Os diplomados das áreas apresentadas, excetuando as Ciências da Saúde (p-value=0.081), têm menos possibilidades de conseguir emprego que os diplomados dos cursos das Engenharias. Comparativamente com os cursos desta área, a área das

Artes e Humanidade regista menos 79% de possibilidades de emprego, seguindo-se a Arquitetura e Construção, com menos 76% de possibilidade e as Ciências Físicas com menos 75%;

- As variáveis género, idade (categorizada), participação em estágios não curriculares não assumem relevo neste modelo logístico.

5. Conclusão

O instrumento utilizado no estudo da trajetória acadêmica e profissional dos diplomados da Instituição de Ensino Superior considerada permite retirar bastante informação (ver anexo). No entanto, a forma como foi desenhado impôs uma análise mais descritiva dos dados recolhidos, onde o objetivo foi identificar e quantificar o grau da relação entre variáveis através da análise de tabelas de contingência. Uma vez identificadas algumas das variáveis que poderiam estar relacionadas com a situação profissional dos diplomados, recorreu-se à análise de regressão logística para a construção de um modelo que permita compreender a empregabilidade através de um conjunto de variáveis.

Este trabalho corrobora a ideia de que efetivamente existem variáveis que podem influenciar a obtenção de emprego, quer o tempo quer o tipo emprego. Na análise das tabelas de contingência ficou patente que o ano de graduação, o ciclo de estudos e a área científica do curso relacionam-se de alguma forma com a situação profissional. Entre as constatações retiradas desta análise destacam-se as seguintes:

- A taxa de emprego é superior entre os diplomados de mestrados (integrados ou não) do que entre os diplomados de licenciatura;
- Os diplomados de licenciatura em média demoram mais tempo na obtenção do primeiro emprego (ainda a diferença seja pouco relevante);
- Na comparação dos três ciclos de estudo, são os licenciados que registam a maior percentagem a trabalhar em profissões que não são da área do seu curso e não requer formação superior;
- O desemprego atinge mais as mulheres;
- A situação ocupacional no último ano do curso relaciona-se com a situação profissional, onde os diplomados que acumularam os estudos com um emprego registam valores inferiores de desemprego;
- As áreas das Ciências Sociais e Direito e das Artes e Humanidade registam as maiores taxas de desemprego;

- A fraca oferta de empregos na área científica é apontada como a principal dificuldade na obtenção de emprego, seguindo-se o excesso de diplomados na área científica.

Por sua vez na análise de regressão logística permitiu confirmar muitas destas conclusões retiradas da análise de tabelas de contingência, de resto esta análise serviu de ponto de partida para a construção do modelo permitindo identificar algumas das variáveis que poderiam contribuir para a empregabilidade.

Assim, inicialmente consideraram-se como candidatas a incluir no modelo logístico as variáveis género, ciclo de estudos, área científica do curso, idade, participação em programas de intercâmbio ou mobilidade, participação em estágios não curriculares, prossecução de estudos e a situação ocupacional no último ano do curso. Porém, a análise de regressão simples de cada variável levou concluir que as variáveis idade e participação em estágios não curriculares não contribuem de forma significativa no valor da variável dependente, e portanto, estas variáveis não foram consideradas na análise de regressão múltipla que permitiu revelar que também o género não influencia o comportamento da empregabilidade.

Após a aplicação da análise foi possível constatar que:

- O modelo obtido inclui as variáveis ciclo de estudos, área científica, a participação em programas de mobilidade, a prossecução de estudos e a situação ocupacional no último ano de curso;
- As variáveis género, idade e participação em estágios não curriculares não contribuem de forma significativa para o modelo da empregabilidade;
- Os diplomados de mestrado integrado têm maiores possibilidades de obter emprego do que os diplomados de licenciaturas (mais 65%);
- Os diplomados de algumas das áreas científicas, em particular das Artes e Humanidades, da Arquitetura e Construção e das Ciências Físicas têm menor hipóteses de obter emprego que os diplomados de curso de Engenharias¹¹.

Na comparação das duas abordagens de análise é possível concluir que existem pontos convergentes. A empregabilidade é influenciada pela área científica do curso, pelo ciclo de estudos e situação ocupacional no último ano. Em contrapartida, na análise das tabelas de

¹¹ Relembre-se que se exclui desta área a engenharia civil, que está na Arquitetura e Construção.

contingência concluiu-se que o género estava relacionado com a situação profissional, facto não corroborado pela análise de regressão logística.

Este trabalho permite concluir que ambas as abordagens conduzem a resultados semelhantes, até porque a metodologia associada é muito similar (ambas recorrem às diferenças entre as frequências esperadas e observadas). Porém, a análise de regressão, além de possibilitar retirar conclusões das variáveis individualmente, tem a vantagem de resultar num modelo que permite prever a empregabilidade.

Seria igualmente importante num trabalho futuro, considerar cada área e cada ciclo de estudos como variáveis dicotómicas. Nesta abordagem teria todo o interesse, por um lado estudar a interação entre as variáveis - por exemplo, a área das engenharias e o ciclo de estudo; por outro lado, prever Y para uma situação particular de interesse - por exemplo, um indivíduo do sexo masculino, formado em engenharia com mestrado integrado e que tenha participado em estágios não curriculares. Neste trabalho foram identificadas algumas das variáveis associadas à empregabilidade e com base na opinião dos diplomados. No entanto, há a consciência que existem outras variáveis que poderão ser igualmente determinantes na obtenção de emprego, nomeadamente, a região ou a Instituição de Ensino. Noutra perspetiva, seria igualmente importante a auscultação aos empregadores, de modo a perceber quais as variáveis que consideram pertinentes num candidato a emprego. Será que coincidem com aquelas que foram obtidas neste estudo? Talvez em trabalhos futuros se consiga responder a esta questão.

Bibliografia

- AGRESTI, Alan: *Categorical Data Analysis*. New Jersey: John Wiley & Sons, Inc., 2002.
- ARCHER, Kellie J.; LEMESHOW, Stanley - *Goodness-of-fit test for a logistic regression model fitted using survey sample data*. The Stata Journal vol. 6, n. 1, 2006, p. 97–105.
- BISHOP, Yvonne M.; HOLLAND, Paul W.; FIENBERG, Stephen E. - *Discrete Multivariate Analysis: theory and practice*. New York: Springer-Verlag, 2007.
- CRAMÉR, Harald - *Mathematical Methods of Statistics*. Princeton, Princeton Univ. Press, 1946.
- EVERIT, Brian S. - *The Analysis of Contingency Tables*. New York: John Wiley & Sons, Inc., 1997.
- FERREIRA, Maria C. C. S. - *Modelos de Regressão: uma aplicação em Medicina Dentária*. Lisboa: Universidade Aberta, 2013. Tese de Mestrado,
- GOODMAN, Leo A.; KRUSKAL, William H. - *Measures of Association for Cross Classification*. IN Journal of the American Statistical Association, vol. 49, n. 268, 1954, p. 732-764.
- HAUCK, WALTER W.; DONNER, Allan - *Wald's test as applied to hypotheses in logit analysis*. IN Journal of the American Statistical Association, vol. 72, 1977, p. 851–853.
- HOSMER, Davis; LEMESHOW, Stanley - *Applied Logistic Regression* (3rd ed). Hoboken, New Jersey: John Wiley & Sons, Inc., 2013.
- KISLAYA, Irina - *Statistical multivariate approaches for identification of predictors of academic failure for first year students in medical school*. Universidade do Minho, 2012. Tese de Mestrado.
- McCULLAGH, Peter; NELDER, John A. - *Generalized Linear Models*, Second Edition. London: Chapman & Hall (1989).
- MINUSSI, João Alberto; DAMACENA, Cláudio; NESS Jr, Walter Lee (2002) - Um modelo de previsão de solvência utilizando regressão logística. IN Rev. Adm. Contemp.[online], vol.6, n.3, 2002, p. 109-128. Disponível em <http://dx.doi.org/10.1590/S1415-65552002000300007>.

NETO, P. L. C. - *Estatística*. São Paulo: Edgard Blucher (16ª reimpressão.), 1977 ou 1998.

PESTANA, Maria H.; GAGEIRO, João N. - *Análise Categórica, Árvores de Decisão e Análise de Conteúdo*. Lisboa: Lidel, 2009.

SAHA, Goutam - *Applying Logistic Regression Model to the Examination Results Data*. Journal of Reliability and Statistical Studies, vol. 4, n. 2, 2011, p. 105-117.

SILVA, Anabela Costa - *Análise Estatística de Inquéritos online*. Universidade do Minho, 2011. Tese de Mestrado.

VENTICINQUE, Eduardo M.; Carneiro, Juliana S.; Moreira, Marcelo P.; Ferreira, Leandro - O uso de regressão logística para espacialização de probabilidades. Megadiversidade, Vol. 3, nº 12, 2007.

[1] <http://www.portalection.com.br/content/an%C3%A1lise-de-regress%C3%A3o>, consultado em setembro de 2014.

Anexo: Questionário

Nota: O questionário apresentado neste anexo refere-se ao que foi utilizado na inquirição dos diplomados de Mestrados Integrados, no entanto, os questionários usados na auscultação dos diplomados de Licenciaturas e de Mestrados, apenas diferem na identificação do curso.

INQUÉRITO SOBRE A TRAJETÓRIA ACADÉMICA E PROFISSIONAL DOS DIPLOMADOS - INQUÉRITO MESTRADO INTEGRADO

A qualidade do ensino e da formação ministrada é uma das principais preocupações desta Instituição e o conhecimento do percurso profissional dos seus diplomados bem como das necessidades de formação ao longo da vida constitui uma informação relevante na aposta que a Instituição faz de ser uma Universidade de referência.

Por essa razão lhe pedimos que preencha este inquérito, com regularidade. Após o 1º preenchimento, as respostas dadas são assumidas por defeito, tendo apenas que atualizar a informação que tenha sofrido alteração desde a sua anterior resposta.

Muito obrigada pelos minutos que vai dedicar ao preenchimento deste inquérito.

IDENTIFICAÇÃO

Data nascimento: ano/mês/dia

Sexo: F / M

Estado Civil: solteiro, casado/em união de facto, separado/divorciado, viúvo

Naturalidade: Portugal, Concelho (abrir para lista); Estrangeiro (abrir para lista de Países)

Residência na altura da candidatura: Portugal, Concelho (abrir para lista); Estrangeiro (abrir para lista de Países)

Residência atual: Portugal, Concelho (abrir para lista); Estrangeiro (abrir para lista de Países)

PERCURSO ACADÉMICO

Nesta Instituição, concluiu o curso de Mestrado Integrado _____

1. Este foi o seu 1.º curso no ensino superior?

Sim

Não, já tinha outro curso concluído nesta Instituição

abrir para licenciatura, mestrado integrado, mestrado, doutoramento

Não já tinha outro(s) curso(s) concluído(s) noutra(s) IES nacional(ais)

Não, já tinha outro(s) curso(s) superior(es) concluído(s) no estrangeiro

2. Relativamente ao curso de mestrado integrado concluído nesta Instituição (a partir daqui considere sempre este curso mesmo que tenha feito outra formação nesta Instituição)

Iniciou em abrir para lista de anos

Terminou em abrir para lista de anos

3. Durante o seu curso participou em algum programa de intercâmbio/mobilidade de estudantes?

Sim

Não

Se sim, indique:

Duração: _____ (número de meses)

País: abrir para lista de países

4. No último ano do curso de mestrado integrado, qual era a sua situação ocupacional?

Só estudava

Estudava e realizava trabalho ocasional

Estudava e trabalhava a tempo completo ou integral

Estudava e trabalhava a tempo parcial

5. Após a conclusão do curso de mestrado integrado prosseguiu estudos no ensino superior (outro curso conferente de grau)?

Sim, nesta Instituição abrir para curso de licenciatura, mestrado integrado,

mestrado , doutoramento

passa à 5.1

Sim, noutra IES em Portugal

passa à 5.1

Sim, noutra IES no estrangeiro

passa à 5.1

Não

passa à 5.3

5.1 Inscreveu-se nesse novo curso (escala *Likert* 5 pontos: discordo completamente, discordo, não concordo nem discordo, concordo, concordo completamente)

Para aprofundar conhecimentos e competências

Para desenvolver competências pessoais (e.g., empreendedorismo, comunicação, trabalho em equipa, resolução de problemas)

Por considerar insuficiente a formação científica anteriormente obtida

Por influência de outros (família, colegas ou professores)

Para continuar a ter benefícios como estudante

Porque não tinha emprego

Para aumentar as oportunidades de emprego

5.2. Relativamente a este novo curso

Ainda prossegue os estudos nesta Instituição

passa à 6

Já concluiu o curso

passa à 6

Desistiu

passa à 6

5.3. Não prosseguiu estudos por (escala *Likert* 5 pontos: discordo completamente, discordo, não concordo nem discordo, concordo, concordo completamente)

Falta de recursos económicos

Ter um emprego em vista

Considerar suficiente a formação dada pelo mestrado integrado

Não pretende continuar a estudar

Por influência de outros (família, colegas ou professores)

passa à 6

PRIMEIRO EMPREGO

6. Após a conclusão do mestrado integrado

Manteve o emprego regular que tinha quando estudava

passa à 7

Manteve o empregado que tinha quando estudava mas mudou de atividade ou categoria

passa à 7

Obteve emprego no local onde realizou estágio curricular

passa à 7

Obteve emprego após _____ (número de meses)

passa à 7

Ainda está à procura de emprego regular

passa à 27

7. Esse emprego situa-se/situava-se em Portugal, Concelho (abrir para lista); Estrangeiro (abrir para lista de Países)

8. Este emprego é ou era na área científica do seu mestrado integrado?

Sim

passa à 9

Não, mas requer formação superior

passa à 20

Não e não requer formação superior

passa à 20

9. Como obteve o 1º emprego na área científica do seu curso de mestrado integrado? (escolha uma opção)

Por anúncio

Por concurso público

Através do Portal/Gabinete de Inserção Profissional desta Instituição

Através do Centro de Emprego

Através da *internet* ou das redes sociais

Através de familiares ou amigos

Através de colegas de curso

Através de professores do curso/faculdade que frequentou

Na sequência de um estágio profissional/curricular

Foi uma autoproposta

Através da criação de uma empresa

Teve uma bolsa de investigação

Outra situação abrir para qual _____

10. Sentiu dificuldades na obtenção deste 1º emprego?

Sim

passa à 10.1

Não

passa à 11

10.1. As dificuldades sentidas na obtenção desse 1º emprego deveram-se

(escala *Likert* 5 pontos: discordo completamente, discordo, não concordo nem discordo, concordo, concordo completamente)

Falta de experiência profissional

Frac. oferta de empregos nesta área científica

Excesso de licenciados nesta área científica

Indisponibilidade para me afastar da residência habitual

11. Indique em que medida a formação obtida nesta Instituição foi importante para a obtenção desse emprego (escala *Likert* de 5 pontos: Sem importância, Pouco importante, Indiferente, Importante, Muito importante)

Conhecimentos teóricos

Conhecimentos técnicos
Competências profissionais
Competências relacionais
Capacidade de adaptação à mudança
Capacidade de resolução de novos problemas
Prestígio da própria universidade

12. Nesse emprego qual é, ou era, a sua situação laboral? (escolha uma opção)

Trabalhador por conta própria com empregados
Trabalhador por conta própria sem empregados
Trabalhador independente abrir para prestação de serviços/recibos verdes/ avença
Trabalhador por conta de outrem abrir para contrato sem termo ou efetivo, a termo, a termo incerto
Bolseir@ num projeto de investigação científica
Está a realizar estágio profissional remunerado
Outra situação abrir para qual _____

13. Qual o rendimento líquido mensal que auferi ou auferia (em euros)?

Igual ou inferior a 500
De 501 – 800
De 801 – 1100
De 1101 – 1400
De 1400 – 1700
Igual ou superior a 1701
N/R

14. Qual o tipo de empresa ou organização onde tem ou teve esse emprego? (escolha uma opção)

Empresa privada
Empresa pública
Empresa mista (capitais públicos e privados)
Órgão da Administração Pública Central e Regional (escolas, tribunais, ministérios, etc.)
Órgão da Administração Local (câmaras municipais, juntas de freguesias)
Instituto Público
Instituição Particular de Solidariedade Social
Outro abrir para qual _____

15. Atualmente, mantém este seu 1º emprego regular?

Sim

passa à 26

Não

passa à 16

16. Quanto tempo permaneceu neste emprego regular?

Menos de 3 meses

Entre 3 e 6 meses

Entre 6 e 12 meses

Entre 12 e 36 meses

Mais de 36 meses

17. Cessou este emprego regular por (escolha uma opção)

Cessaç o do contrato

passa à 19

Despedimento por iniciativa da entidade empregadora

passa à 19

Despedimento por iniciativa pr pria

passa à 18

Rescis o contratual por m tuo acordo

passa à 19

Fal ncia da empresa

passa à 19

Cessaç o da bolsa de investiga  o

passa à 19

Outra abrir para qual _____

passa à 19

18. Decidiu cessar este emprego por pretender ou lhe ser oferecida (escala *Likert* 5 pontos:

discordo completamente, discordo, n o concordo nem discordo, concordo, concordo completamente)

Melhor remunera  o

Melhores condi   es de trabalho

Aproxima  o da zona de resid ncia

Projeto de trabalho mais aliciante

Institui  o ou fun   o de maior prest gio

Raz  es de sa de ou familiares

19. Desde que deixou o seu 1  emprego regular esteve desempregad@

Uma vez

Duas vezes ou mais vezes (abrir para: por per odos inferiores a 6 meses; por per odos superiores a 6 meses)

OUTRA SITUAÇÃO PROFISSIONAL

20. Qual é a sua situação atual?

Empregad@ (a tempo inteiro ou parcial)

passa à 21

Desempregad@ à procura de novo emprego regular

passa à 27

Bolseir@ de investigação

passa à 22

A realizar estágio profissional remunerado

passa à 22

A realizar estágio profissional não remunerado

passa à 28

Estudante, a frequentar um curso de

abrir para licenciatura, mestrado, doutoramento

abrir para a Instituição ou outra IES

passa à 30

Outra situação abrir para qual _____

passa à 21

21. Qual é a sua condição laboral (assinale apenas uma resposta)?

Trabalhador por conta própria com empregados

Trabalhador por conta própria sem empregados

Trabalhador independente abrir para prestação de serviços/recibos verdes/ avença

Trabalhador por conta de outrem abrir para contrato sem termo ou efetivo, a termo, a termo incerto

22. Que rendimento líquido mensal que auferes (em euros)?

Igual ou inferior a 500

De 501 – 800

De 801 – 1100

De 1101 – 1400

De 1400 – 1700

Igual ou superior a 1701

N/R

23. Em que tipo de empresa ou organização trabalha? (escolha uma opção)

Empresa privada

Empresa pública

Empresa mista (capitais públicos e privados)

Órgão da Administração Pública Central e Regional (escolas, tribunais, ministérios, etc.)

Órgão da Administração Local (câmaras municipais, juntas de freguesias)

Instituto Público

Instituição Particular de Solidariedade Social

Outro abrir para qual _____

24. Desenvolve o seu trabalho na área científica da sua formação académica

Sim

Não, mas o meu emprego atual requer formação superior

Não e o meu emprego atual não requer formação superior

25. Este emprego situa-se Portugal, Concelho (abrir para lista); **Estrangeiro** (abrir para lista de Países)

26. Atualmente, em termos laborais, pretende (escolha uma opção)

Manter a situação ocupacional atual

Permanecer na empresa ou organização mas progredir na carreira

Mudar de empresa ou organização

Mudar de empresa ou organização e mudar de atividade profissional

Outra abrir para qual _____

passa à 28

27. As dificuldades sentidas na obtenção de emprego devem-se (escala *Likert* 5 pontos: discordo completamente, discordo, não concordo nem discordo, concordo, concordo completamente)

Falta de experiência profissional

Fracas oferta de empregos nesta área científica

Excesso de licenciados nesta área científica

Indisponibilidade para me afastar da residência habitual

Más condições remuneratórias de outras propostas

28. Atualmente, pretende realizar nova formação nesta Instituição?

Sim, um curso conferente de grau

abrir para licenciatura, mestrado, doutoramento

passa à 30

Sim, uma formação não conferente de grau

passa à 30

Não

passa à 29

29. Não pretendo fazer atualmente nenhuma formação nesta Instituição porque

Não existe nenhum curso na área que pretendo

abrir para caixa de texto em que é referida a área

Pretendo conhecer novos docentes e/ou novas perspetivas

Pretendo aproximar-me da área de residência e/ou de trabalho

Não pretendo fazer qualquer nova formação

30. Pretende receber informação sobre iniciativas e apoios na área do empreendedorismo?

31. Gostaria de ser contactado para conhecer melhor a oferta formativa desta Instituição?

Sim

indique a melhor forma de o contactar (e-mail ; telefone)

Não

Muito obrigada pela sua participação. Submeter